Title: **In defense of Apocynaceae: inference on evolution of pyrrolizidine alkaloids from evolution of an enzyme in their biosynthetic pathway, homospermidine synthase.**

Authors: Chelsea R. Smith[1,2], Elisabeth Kaltenegger[3], Jordan Teisher[1,4], Abigail J. Moore[5], Shannon C. K. Straub[6], Tatyana Livshultz[1,2]*

Affiliations:

[1]Botany Department, Academy of Natural Sciences of Drexel University, Philadelphia, PA, USA

[2]Department of Biodiversity, Earth, and Environmental Sciences, Drexel University, Philadelphia, PA, USA

[3]Botanisches Institut und Botanischer Garten, Christian-Albrechts-Universität zu Kiel, Kiel, Germany

[4]MO Herbarium, Missouri Botanical Garden, St. Louis, MO, USA

[5]Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK, USA

[6]Department of Biology, Hobart and William Smith Colleges, Geneva, NY, USA

*Correspondence to: tl534@drexel.edu

Abstract (249 words): If plant defenses experience positive selection when they first evolve, evidence of selection on a gene can place the origin of novel defenses on a phylogeny. Prior research on pyrrolizidine alkaloids (PAs), rare specialized metabolites of Apocynaceae, reconstructed evolution of homospermidine synthase (HSS), an enzyme of PA biosynthesis, to infer a single origin of PAs by showing a single duplication in the MRCA of all known PA-producing species that gave rise to an enzyme with the predicted motif (VXXXD) of a

functionally optimized HSS. We follow up by testing the effect of amino acid motif on HSS function, revisiting motif evolution in *HSS*-like genes, and testing for selection to infer evolution of HSS function. Some evidence supports a single origin of PAs: an IXXXD HSS, similar in function to VXXXD HSS, evolved in the MRCA of all PA-producing species; loss of optimized HSS function occurred multiple times via pseudogenization and perhaps via evolution of an IXXXN motif. Other evidence indicates multiple origins: the VXXXD motif, present in all PA-producing species, evolved two or three times independently; branch-site models suggest that the ancestral IXXXD branch was not under positive selection while some VXXXD branches were; there is no evidence of relaxed selection on putatively impaired *HSS*-like genes (IXXXN motif); substitutions at sites experiencing positive selection occurred on multiple branches in the *HSS*-like gene tree, suggesting that HSS function may have been optimized in multiple independent lineages. We discuss how complexity of genotype-phenotype maps confounds inference of phenotypic from genic evolution.

Key words

Apocynaceae, homospermidine synthase, deoxyhypusine synthase, pyrrolizidine alkaloids, duplication-degeneration-complementation model, plant defense, functional evolution, enzyme evolution, specialized metabolism

Introduction

Adaptation of organismal phenotypes to their environments by natural selection is one of the central predictions of the theory of evolution (Darwin 1859). Yet, because natural selection acts within populations, testing adaptive hypotheses about phenotypes that are fixed within species is

a central challenge in evolutionary biology (Reeve and Sherman 1993; Olson-Manning, et al. 2012). The neutral theory of molecular evolution (Kimura 1968b; Kimura 1968a) has permitted inference of phenotypic adaptation via reconstruction of the evolutionary dynamics of the genes that construct ecologically significant phenotypes (Kroymann 2011; Joly-Lopez, et al. 2016; Scossa and Fernie 2020). Specifically, signatures of positive selection in gene evolution provide evidence of adaptive evolution of novel phenotypes, for example in Amaranthaceae *rbcL* during the transition from C_3 to C_4 photosynthesis (Kapralov, et al. 2012; Parto and Lartillot 2018).(Parto and Lartillot 2018) Purifying selection on genes is evidence of selective maintenance of a functional optimum, as seen in the diatom silicon transporter (SIT) gene family that facilitates building of the diatom cell wall (Alverson 2007). Neutral genic evolution is evidence of selective neutrality of a phenotype, as seen in Orobanchaceae plastomes as lineages shifted to an obligate parasitic lifestyle (Wicke, et al. 2016), or it may be evidence of selection against a phenotype (Cutter and Jovelin 2015), as in carnivorous polar bears, where there was diet-related gene loss, as traits associated with an omnivorous diet experienced negative selection (Rinker, et al. 2019).

Plant specialized metabolism, production of non-essential small molecules (Pichersky and Lewinsohn 2011; Moghe and Last 2015), is an ideal system for study of linked adaptive evolution of genotype and phenotype (Benderoth, et al. 2006; Kroymann 2011; Rausher and Huang 2016; Feng, et al. 2019). Plant specialized metabolites mediate interactions with the environment, including defense against herbivores, and thus likely have large fitness consequences for the plant and multiple opportunities for selection (Moore, et al. 2014). Novel toxic metabolites are predicted to increase plant fitness by deterring unadapted herbivores (Ali and Agrawal 2012), and therefore the enzymes involved in the biosynthesis of toxic metabolites

are predicted to experience strong positive selection (Scossa and Fernie 2020). However, adapted specialist herbivores can co-opt metabolites at the expense of their host plant (Ehrlich and Raven 1964; Erb and Robert 2016; Erb and Reymond 2019), resulting in strong selection against the production of the co-opted metabolites (Agrawal, et al. 2009; Cacho, et al. 2015). The great diversity of plant specialized metabolites (Wink 2011) is the result of the functional divergence of plant biosynthetic enzymes (Pichersky and Gang 2000; Pichersky and Lewinsohn 2011), encoded by gene families which have evolved via gene duplication, providing abundant raw materials for selection for novel regulatory and catalytic functions (Ober 2010; Scossa and Fernie 2020). For example, genes of the terpene synthase (TPS) family evolved to use multiple substrates and have diverged in tissue-specific expression in *Arabidopsis* (Chen, et al. 2004), and the first three steps in caffeine biosynthesis in *Coffea* are catalyzed by gene duplicates (Denoeud, et al. 2014).

Pyrrolizidine alkaloids (PAs) are a group of plant specialized metabolites characterized by a necine base, derived from homospermidine, with one or more attached necic acids (Hartmann and Witte 1995; Ober and Hartmann 2000). PAs likely evolved as a defense against insect herbivores (Hartmann 1999; Ober and Kaltenegger 2009) and occur in 16 angiosperm families (Hartmann and Witte 1995; Nawaz, et al. 2000; Kim, et al. 2010; Tamariz, et al. 2018; Barny, et al. 2021). Homospermidine synthase (HSS), the first enzyme in the PA biosynthetic pathway, evolved via gene duplication from deoxyhypusine synthase (DHS) (Ober and Hartmann 2000), a highly conserved enzyme across eukaryotes that activates eIF5A, a translation elongation factor (Chen and Liu 1997; Park, et al. 2010) (Fig. 1). HSS has lost this ancestral DHS function, but has retained and optimized another function of DHS: the ability to catalyze the reaction of putrescine with spermidine, producing an order of magnitude more

homospermidine than DHS (Ober, Gibas, et al. 2003; Reimann, et al. 2004; Kaltenegger, et al. 2013). Eight independent *DHS/HSS* duplications occurred in seven PA-producing angiosperm families (Anke, et al. 2004; Reimann, et al. 2004; Kaltenegger, et al. 2013; Gill, et al. 2018; Livshultz, et al. 2018), suggesting that PA biosynthesis evolved repeatedly.

Kaltenegger et al. (2013) used branch-site models to identify three amino acid sites that are under positive selection in the evolution of HSS from DHS in Convolvulaceae: N266H, I277V, and N281D (Kaltenegger, et al. 2013). Mutagenesis experiments on *Ipomoea neei* DHS supported the functional importance of these amino acid substitutions. The modified DHS had reduced eIF5A activation activity, similar to HSS (Kaltenegger, et al. 2013). Two of the three amino acid changes associated with the evolution of HSS (VXXXD) from DHS (IXXXN) in Convolvulaceae have evolved independently in the functionally characterized *HSS* loci of each of 7 plant families (Livshultz, et al. 2018), suggesting that the VXXXD motif is important to optimization of homospermidine production or loss of eIF5A activation.

Livshultz et al. (2018) showed that the pattern of *DHS/HSS* duplication, IXXXN/VXXXD motif evolution, and pseudogenization in Apocynaceae supports predictions of the defense de-escalation hypothesis: that PA biosynthesis evolved once in Apocynaceae and was lost in multiple lineages due to negative selection from adapted herbivores (Edgar 1984). Livshultz et al. (2018) identified a single *DHS/HSS* duplication in the most recent common ancestor (MRCA) of all PA-producing taxa in Apocynaceae, early in the diversification of the APSA clade (Fig. 2),and ancestral character state reconstruction supported an ancestral HSS with a VXXXD motif. HSS enzymes in genera known for PAs [*Parsonsia, Prestonia, Echites* (Echiteae)*, Anodendron* (Apocyneae)] retain the VXXXD motif with one exception [*Alafia* (Nerieae), IXXXD) while genera not known for PAs either retained the VXXXD motif or

evolved IXXXD or IXXXN motifs (Figs. 2, 3; Barny, et al. 2021). Furthermore, *HSS* is pseudogenized in *Asclepias syriaca* (Asclepiadeae), a species which produces cardenolides but no PAs (Livshultz, et al. 2018) (Fig. 2).

If the origin of pyrrolizidine alkaloids was an adaptation, and the optimization of HSS function occurred concurrently with assembly of the PA biosynthetic pathway, we can make testable predictions about patterns of selection that we should observe on *HSS*-like genes in Apocynaceae. If there was a single origin of PAs in the MRCA of all PA-producing taxa (Figs. 3a, c), we predict positive selection on the *HSS*-like gene of this MRCA followed by purifying selection on branches that retained optimized HSS function (i.e. VXXXD motif or PA phenotype) and relaxed selection on branches that lost optimized HSS function (e.g., pseudogenization or evolution of IXXXN motif). In contrast, if parallel recruitment of the ancestral *HSS*-like paralog into independently evolved PA biosynthetic pathways occurred (Figs. 3b, d), we would predict either purifying (if DHS function was retained) or relaxed (if DHS function was lost) selection on the ancestral *HSS*-like gene followed by positive selection on the ancestral *HSS* of each lineage were the HSS VXXXD motif and/or PA-production evolved. Under this scenario, branches with the IXXXN motif should be under purifying selection (if they retained DHS function) or under relaxed selection (if they lost DHS function).

In the present study, we revisit the evolution of *HSS*-like genes in Apocynaceae to test additional predictions of the defense de-escalation hypothesis. We again reconstruct ancestral motifs and ask if patterns of motif evolution and ratios of synonymous to non-synonymous amino acid substitutions (ω) during the evolution of *HSS*-like genes are consistent with a single origin of optimized HSS function (Figs. 3a, c) or with multiple independent origins (Figs. 3b, d). We further ask if there is evidence that the HSS VXXXD motif is under positive selection in

Apocynaceae as has been previously demonstrated in Convolvulaceae (Kaltenegger, et al. 2013), and if any other amino acid substitutions were selected during the evolution of optimized HSS function. We also investigate if the VXXXD motif is important to HSS function in Apocynaceae. Finally, we consider how optimization of gene function, as evidenced by evolution of a functionally important motif or positive selection, is pertinent to locating the origin of a novel biosynthetic pathway on a phylogenetic tree.

Results

Taxon sampling

We obtained *DHS*-like and *HSS*-like sequences from 142 species in 14 of the 16 APSA clade tribes or subfamilies (Apocyneae, Asclepiadeae, Baisseeae, Echiteae, Fockeeae, Malouetieae, Marsdenieae, Mesechiteae, Nerieae, Odontadenieae, Periplocoideae, Rhabdadenieae, Secamonoideae, and Wrightieae; not sampled: Ceropegieae, Eustegieae) and 9 of 11 rauvolfioid tribes (Alyxieae, Amsonieae, Aspidospermateae, Carisseae, Hunterieae, Melodineae, Plumerieae, Tabernaemontaneae, Vinceae; not sampled: Alstonieae, Willughbeieae) (Fig. 2), taxonomy according to Endress, et al. (2018), from new sequencing and published datasets (Hoopes, et al. 2018; Livshultz, et al. 2018). *Gelsemium sempervirens* (Gelsemiaceae) was included as an outgroup (Table S1).

Assembly and annotation of DHS- and HSS-like loci

Our automated assembly pipeline yielded 437 annotatable contigs (Table S1) from 122 libraries. Often exons 1 and 7 were more than 50% divergent from the reference *Parsonsia alboflavescens* sequences and required manual annotation. One hundred thirteen (113) of the 437 contigs contained a complete CDS covering exons 1 to 7. The exon sequences from all annotated

contigs were extracted and aligned with the previously published Apocynaceae *DHS*/*HSS* dataset (Livshultz, et al. 2018) and sequences recovered from the *Calotropis gigantea* genome (initial alignment, Table S2) to construct an initial gene tree (Fig. S1). Contigs with a partial ORF were then manually concatenated according to a topology-aware algorithm (see 'gene assembly from contigs' in Materials & Methods).

Orthologs of *P. alboflavescens HSS* (84% bootstrap support [BS]) were designated as *HSS*-like genes; all other sequences were designated as *DHS*-like (Fig. S1, Table S1). All rauvolfioid species (20 libraries), all species of Wrightieae (5 libraries), and *Gelsemium sempervirens* (1 library) yielded only *DHS*-like contigs. All APSA clade species (96 libraries) except Wrightieae had both *DHS*- and *HSS*-like contigs.

After concatenation, we recovered at least one *DHS*-like gene with an intact ORF covering at least exons 2 through 6 from all 122 libraries (Table S1). Of the 60 libraries that did not require manual concatenation of *DHS*-like contigs, 11 had additional short contigs, and 7 had 2 or 3 complete *DHS*-like ORFs, either alleles or paralogs. The remaining 62 libraries each yielded 2-7 partial *DHS*-like contigs that were combined to form complete ORFs. Typically, the partial contigs covered exons 1-4 and exons 5-7, likely due to low probe density in intron 4. There is at least one *HSS*-like gene with an intact ORF covering at least exons 2 through 6 from each of 87 libraries, 71 of which did not require concatenation. Twenty (20) of these 71 have additional short contigs, and 15 have 2 to 5 complete *HSS*-like ORFs, either alleles or paralogs. For 16 libraries, a complete *HSS*-like ORF was assembled from 2-5 partial contigs (Table S1).

Validation of new sequencing and assembly

Twenty-nine (29) accessions previously sequenced by Livshultz et al. (2018) with the Sanger method were re-sequenced and assembled (Table S1). All 36 re-sequenced loci have over

96% amino acid identity in the region of overlap with the Sanger sequences. We also found 43 additional sequences from these 29 libraries: full *DHS*-like sequences in 11 libraries, full *HSS*-like sequences in 6 libraries, short *DHS*- or *HSS*-like sequences in six libraries, and *DHS*- or *HSS*-like pseudogene sequences in 8 libraries.

Characteristics of DHS- and HSS-like sequences

Within a genome, similarity between complete or nearly complete ORFs of *DHS*- and *HSS*-like paralogs ranged from 72.2% (*Sarcolobus cambogensis*) to 89.4% (*Malouetiella mildbraedii*) amino acid identity, less than among paralogous or allelic *DHS*-like ORFs 89.7% (*Pleioceras barteri*, *Wrightia lanceolata*) to 100% (*Echites umbellatus*, *Finlaysonia insularum*) or among *HSS*-like ORFs 90.3% (*Gymnema maingayi*) to 99.6% (*Gongronemopsis angolense*) in the same genome (Table S4). Across species, pairwise similarity among complete *HSS*-like ORFs ranged from 77.1% (*Dischidia albida* and *Isonema smeathmannii*) to 99.1% (*Aganosma schlechteriana* and *Epigynum cochinchinensis*). Similarity between complete *DHS*-like ORFs ranged from 86.6% (*Baissea myrtifolia* and *Allamanda schottii*) to 99.7% (*Tabernaemontana pandacaqui* and *T. peduncularis*) (Table S4). This suggests that *HSS* has diverged more than *DHS*: minimum similarity between species, 77.1% among *HSS*-like genes versus 86.6% among *DHS*-like genes.

Both intact *DHS*-like genes and *DHS*-like pseudogenes occur in four species of three rauvolfioid tribes: Amsonieae (*Amsonia orientalis*), Melodineae (*Diplorhynchus condylocarpon*), and Vinceae (*Ochrosia coccinea, O. poweri*), and four species of four APSA clade lineages: Nerieae (*Strophanthus boivinii*), Baisseeae (*Baissea myrtifolia*), Apocyneae (*Trachelospermum axillare*), and Echiteae (*Laubertia boissieri*) (Table S1). Both intact *HSS*-like genes and *HSS*-like pseudogenes occur in 14 species of seven APSA clade lineages: Malouetieae

(*Malouetiella mildbraedii, Funtumia elastica*), Periplocoideae (*Raphionacme flanaganii*), Odontadenieae (*Pinochia corymbosa*), Mesechiteae (*Forsteronia guyanensis*), Echiteae (*Parsonsia alboflavescens, Echites turriger, E. umbellatus*), Asclepiadeae (*Calotropis gigantea, Tassadia propinqua*), and Marsdenieae (*Gongronemopsis truncata, Gymnema sylvestre, Leichhardtia coronata, Ruehssia guaranitica*) (Table S1). Eight species in five APSA lineages have only *HSS*-like pseudogenes: Apocyneae (*Apocynum androsaemifolium, A. cannabinum, A. pictum, A. venetum* [two accessions]), Asclepiadeae (*Asclepias syriaca),* Malouetieae (*Pachypodium baronii*), Nerieae (*Nerium oleander, Strophanthus preussii*), and Periplocoideae (*Zygostelma benthamii*) (Fig. 2). No species had both *DHS*-like and *HSS*-like pseudogenes.

The recovered putative pseudogenes range from the complete 1061 bp CDS *DHS*-like pseudogene from *Ochrosia coccinea* to only a single 105 bp *HSS*-like exon 5 from *Raphionacme flanaganii*. The number of distinct *DHS*-like or *HSS*-like candidate pseudogene sequences per genome ranges from one (most species) to seven (*Diplorhynchus condylocarpon*) (Table S1). Pairwise nucleotide sequence similarity of the restored ORF among pseudogenes from the same sample library ranges from 55% to 100% (Table S4). Multiple pseudogenes from six of the seven libraries all form a monophyletic clade; the one exception is a pseudogene from *Ochrosia poweri* which falls outside the clade of Vinceae sequences (Fig. S1). Pseudogenes occur in 32 separate clades in the gene tree (Fig. S2), indicative of an active gene birth/death process.

Distribution of DHS and HSS motifs

All *DHS*-like sequences with intact ORFs encode the expected IXXXN motif at positions 269 and 273, except for *Strophanthus boivinii* (MXXXN) (Table S1). *HSS*-like sequences with intact ORFs encode IXXXN and IXXXD motifs along with the predicted VXXXD motif. Amino acid motifs in translated *HSS*-like sequences are conserved at the tribal/subfamilial level. *HSS*-

like genes in Malouetieae, Apocyneae, Mesechiteae, Odontadenieae, and Echiteae code for a VXXXD motif (Fig. 2). Sequences from Secamonoideae, Periplocoideae, Baisseeae, Fockeeae, and Rhabdadenieae code for IXXXD, and Asclepiadeae sequences code for IXXXN (Fig. 2). Nerieae and Marsdenieae have multiple motifs. Among Nerieae, *Strophanthus* and *Isonema* have VXXXD, while *Alafia* has IXXXD (Figs. 2, 3). Marsdenieae *HSS*-like sequences all code for either IXXXD or IXXXN, with one exception, *Dischidia albida* (MXXXT) (Figs. 2, S4).

Most pseudogenes, when their ORF is restored, have the expected amino acid motif based on their tribal position (Table S1). Of the 10 *DHS*-like pseudogenes with the motif area, seven have the expected IXXXN motif and three do not: one of seven *Diplorhynchus condylocarpon* pseudogenes (LXXXH), *Ochrosia poweri* (IXXXD), and *O. coccinea* (TXXXD). Of the 12 *HSS*-like pseudogenes from tribes with a VXXXD motif, seven have the expected motif and five do not: *Apocynum venetum* (VXXXN and VXXXS), *Pinochia corymbosa* (IXXXD), and *Pachypodium baronii* (IXXXD). All 10 of the *HSS*-like pseudogenes in Periplocoideae (IXXXD), Marsdenieae (IXXXN or IXXXD), and Nerieae (VXXXD or IXXXD) have an IXXXD motif. Within Asclepiadeae (IXXXN), the *Calotropis gigantea HSS*-like pseudogene has IXXXD (Table S1).

VXXXD motif occurs in all PA-producing species

All 10 sequenced Apocynaceae species known to produce PAs (*Amphineurion marginatum*, *Anodendron affine*, *Echites umbellatus*, *Laubertia boissieri*, *Macropharynx peltata*, *Parsonsia alboflavescens*, *P. eucalyptophylla*, *Prestonia portobellensis*, *Rhodocalyx rotundifolius*, *Temnadenia odorifera*) (Burzynski, et al. 2015; Colegate, et al. 2016; Barny, et al. 2021) have an intact *HSS*-like gene that encodes a VXXXD motif (Table S1).

Construction of final alignments

The final alignments contained sequences with an intact ORF covering at least exons 2-6, which includes 120 *HSS*-like and 155 *DHS*-like sequences, 238 of them new (Table S1). The "full dataset" also contained 34 shorter sequences (Table S2), as well as 16 *DHS*-like and 40 *HSS*-like pseudogenes. Short sequences and pseudogenes were excluded in the "reduced dataset" (Table S2). Maximum likelihood trees from these alignments are shown in Figs. 4, and S3 (reduced dataset) and S2 (full dataset). Summary trees are depicted in Fig. 3.

Gene tree topologies

Topologies from the full dataset (Figs. 3c, d, S2, S3) and the reduced dataset (Figs. 3a, b, 4) are very similar. The most important differences are the position of the Wrightieae *DHS*-like 2 clade, sister to the *HSS*-like clade (Branch A) in the reduced dataset tree (Figs. 3a, 3b, 4c) or sister to the APSA *DHS*-like clade in the full dataset tree (Figs. 3c, 3d, S2), and the monophyly of Nerieae plus Malouetieae *HSS*-like genes in the reduced dataset tree (Branch F, Figs. 3a, 3b, 4c) or their paraphyly in the full dataset (Figs. 3c, 3d, S2). These relationships were unsupported by bootstrap, and a Shimodaira-Hasegawa (SH) test showed that the reduced dataset does not reject the alternate placement of the Wrightieae *DHS*-like 2 clade nor paraphyly of Nerieae and Malouetieae *HSS*-like sequences (D(LH)=23.93, $p>0.05$, Table S5). Bootstrap support values are higher in the reduced dataset tree (Fig. 4) relative to the full dataset tree (Fig. S2), likely due to missing data introduced by the pseudogenes and short sequences. Thus, the gene tree topology will be discussed with reference to the reduced dataset tree (Fig. 3c, d, 4) except when noted.

All APSA clade sequences form a highly supported monophyletic clade (98 BS), nested in a grade of rauvolfioid sequences (Fig. 4a). Reconciliation of this gene tree topology with the APSA clade phylogeny inferred based on Angiosperm353 nuclear gene sequences (Fig. 2;

Antonelli, et al. 2021) implies two gene duplications in the MRCA of the APSA clade. The first gave rise to the Wrightieae *DHS*-like 1 clade (93 BS) and the rest of the APSA clade sequences (92 BS). The second gave rise to two unsupported clades: one includes the *HSS*-like clade (100 BS) plus the Wrightieae *DHS*-like 2 clade, the second comprises all other APSA *DHS*-like sequences (Fig. 4b). As mentioned above, the Wrightieae *DHS*-like 2 clade is alternately sister to the other APSA *DHS*-like sequences (Figs. S2, S3).

The distribution of pseudogenes on the full dataset tree (Fig. S2) implies additional gene duplications. Pseudogenes occur in 32 different clades but the exact number of duplications is hard to determine because of absence of support for the positions of most pseudogenes on the tree and uncertainty whether *HSS*-like pseudogenes from genomes that also have intact *HSS*-like genes represent alleles or loci. Some pseudogenes are clearly the result of duplications that predated divergence of species: e.g. *Ochrosia poweri* and *O. coccinea* (Vinceae), intact ORFs (99 BS) and pseudogenes (100 BS), *Funtumia elastica* and *Malouetiella mildbraedii* (*HSS*-like clade Malouetieae), intact ORFs (81 BS) and pseudogenes (98 BS), and *Calotropis gigantea*, whose pseudogene is sister to the remainder of the Asclepiadeae *HSS*-like clade (98 BS) (Fig. S2). Other potential duplications that gave rise to pseudogenes appear to post-date the most recent speciation events captured in our dataset (e.g. *HSS*-like genes of *Tassadia propinqua* and *DHS*-like genes of *Baissea myrtifolia*) (Fig. S2).

Within each of the three primary APSA clade gene lineages (Wrightieae *DHS*-like 2, APSA *DHS*-like, APSA *HSS*-like + Wrightieae *DHS*-like 1), clades of sequences from related taxa (classified in the same genus, tribe, or subfamily) are monophyletic and mostly moderately to well-supported (Figs. 4, S3). Exceptions include tribes and genera whose monophyly has not been consistently supported in previous phylogenetic analyses (Odontadenieae, Nerieae,

Marsdenia) (Livshultz, et al. 2007; Morales, et al. 2017; Fishbein, et al. 2018; Espírito Santo, et al. 2019; Liede-Schumann, et al. 2022) , genera with expanded species sampling relative to previous analyses (*Wrightia* in the Wrightieae *DHS*-like 2 clade), and the position of *Eucorymbia alba* (Malouetieae) as sister to *Amphineurion marginatum* (100 BS *HSS*-like clade, 98 BS *DHS*-like clade) in the Apocyneae. These are the first published sequences from the monotypic *Eucorymbia* and this placement likely reflects its phylogeny. Relationships among tribal and subfamilial gene lineages also reflect species phylogeny (Fig. 2). In both *HSS*-like and *DHS*-like subtrees, Wrightieae, Nerieae, and Malouetieae sequences diverge prior to the "crown clade" radiation comprising the rest of the APSA clade (Straub, et al. 2014; Fishbein, et al. 2018; Antonelli, et al. 2021); Baisseeae are sister to the Secamonoideae and Asclepiadoideae clade (Fockeeae, Asclepiadeae, Marsdenieae) (Antonelli, et al. 2021). Sequences from the predominantly New World Mesechiteae, Odontadenieae, and Echiteae form the "MOE" clade (83 BS *HSS*-like clade, 88 BS *DHS*-like clade) (Livshultz, et al. 2007; Morales, et al. 2017). The incongruent positions of "MOE", Apocyneae, Rhabdadenieae, and Periplocoideae between the *DHS*- and *HSS*-like subtrees likely reflect long branch attraction and/or incomplete lineage sorting in a known rapid radiation (Whitfield and Lockhart 2007; Straub, et al. 2014).

At least one duplication gave rise to the paralogous IXXXN *HSS*-like and IXXXD *HSS*-like loci found in 13 of 23 Marsdenieae species. While the maximum likelihood topology resolves the IXXXD *HSS*-like loci as a grade forming successive sister lineages to an unsupported IXXXN *HSS*-like clade (Clade J; Figs. 3, 4, S3), an SH test cannot reject the alterative topology with the grade-forming IXXXD *HSS*-like sequences instead forming a monophyletic clade [D(LH)=4.56, p>0.05, Table S5.] Motifs are homogenous in these two gene lineages except for a *Marsdenia tinctoria* locus (IXXXD) and *Dischidia albida* (MXXXT) locus

nested in the IXXXN *HSS*-like clade and the *Gongronemopsis truncata* hss2 sequence (IXXXN) in the IXXXD *HSS*-like grade (Fig. S4a). These sequences were assembled by the pipeline without manual concatenation and read mapping did not reveal any obvious assembly errors. To test for recombinant origins of discordant motifs in *M. tinctoria* and *G. truncata* sequences, we located potential recombination breakpoints with GARD (Kosakovsky Pond, et al. 2006) on a matrix comprised of only Marsdenieae *HSS*-like sequences (Table S2). The best fit model (AIC_C=11066.18, ΔAIC_C=1135.64) identified one breakpoint between positions 224 and 225 (within exon 4). Splitting the *M. tinctoria* and *G. truncata* sequences each into 2 terminals and realigning and reanalyzing them with the other Marsdenieae *HSS*-like sequences resulted in a topology with both fragments of the *M. tinctoria* sequences nested in the IXXXN clade (Fig. S4b). The 5' fragment of the *G. truncata HSS*2 sequence is nested within the IXXXN clade rather than with the 3' fragment in the IXXXD grade, suggesting that, although this sequence is recombinant, the discordant placement of its amino acid motif is not a product of recombination.

As with the APSA clade subtree, most monophyletic gene lineages in the rauvolfioid grade reflect tribal phylogeny (Figs. 2, 4a, S3), as does the polyphyly of Melodineae (*Pycnobotrya*, *Diplorhynchus*, *Craspidospermum*) (Fishbein, et al. 2018). Higher level relationships of gene lineages from rauvolfioid tribes are unsupported except for placement of *Haplophyton crooksii* (Aspidospermateae) as sister to a clade that includes all other Apocynaceae sequences (96 BS) (Figs. 4a, S3).

Ancestral HSS-like motif reconstruction

To reconstruct ancestral HSS-like motifs, we reconstructed evolution of the reduced dataset alignment on both the reduced dataset tree topology (Figs. 3a, b, 4) and a topology constrained to match the full dataset topology (Figs. 3c, d, S3) with both joint and marginal

(Tables S6, S7) likelihood in PAML (Yang 1997; Yang 2007). The two methods gave highest likelihood to the same reconstructed motifs, hence the marginal reconstruction will be discussed (Tables S6, S7).

On the full dataset tree topology (Figs. 3c, d, S3), the ancestral motif for the *HSS*-like subtree was IXXXD (Branch A, posterior probabilities: I 0.749, D 0.999, evolving from I 1.0, N 1.0). VXXXD motifs evolved 4 times from IXXXD motifs: 1) *Strophanthus boivinii* from MRCA Nerieae (I 0.684, D 1.0), 2) *Isonema smeathmannii* from MRCA *Isonema*+*Alafia* (I 0.686, D 0.999), 3) MRCA Malouetieae (Branch F: V 0.961, D 1.0) from MRCA of core APSA excluding Nerieae (I 0.748, D 1.0), and 4) MRCA of Apocyneae+MOE (Branch B: V 0.998, D 1.0) from MRCA of Apocyneae+MOE+Rhabdadenieae (I 0.959, D 1.0). There were also three substitutions from IXXXD to IXXXN, a reverse mutation back to the DHS-like motif: 1) in the MRCA of the Asclepiadeae tribe (Branch I: I 1.0, N 0.990) from MRCA of Asclepiadeae+Marsdenieae (I 1.0, D 0.997), 2) in the MRCA of the Marsdenieae IXXXN *HSS*-like clade (Branch J: I 1.0, N 0.999) from the MRCA of Clade J+*Gymnema maingayi HSS*-like 3 (I 1.0, D 0.996), and 3) in *Gongronemopsis truncata HSS*-like 2 from its ancestor (I 1.0, D 1.0). A single re-evolution of IXXXD from IXXXN (I 1.0, N 0.998 at MRCA *Marsdenia glabra HSS*-like+*M. tinctoria HSS*-like) occurred in the *Marsdenia tinctoria HSS*-like gene.

The reduced dataset tree topology (Figs. 3a, b, 4) differed only in that a VXXXD motif evolved twice rather than three times from the ancestral HSS-like IXXXD motif: in the MRCA of Apocyneae+MOE (Branch B) and in the MRCA of Nerieae-Malouetieae (Branch G: V 1.0, D 0.995). From a VXXXD (V 0.914, D 0.999 at MRCA *Isonema*+*Alafia*) there was one reversal to IXXXD in the MRCA of the two *Alafia* species (Branch K: I 1.0, D 1.0).

aBSREL – evidence for origin of optimized HSS and PA biosynthesis from selection on HSS

We used aBSREL (Smith, et al. 2015), which tests if a proportion of sites in specified branches evolved under positive selection, to test our main hypotheses: single (Figs. 3a, c) versus multiple convergent origins (Fig. 3b, d) of optimized HSS and PA biosynthesis. aBSREL fits an optimal number of ω (non-synonymous to synonymous substitution) rate classes (ω_1- ω_3) to each branch to allow for selection heterogeneity (Smith, et al. 2015). We reconstructed ω on the selected branches in the Reduced and Full dataset topologies (Tables 1, S8, S9, S10; Figs. 3, 4, S3). Selection criteria for tested branches are i) evolution toward an HSS-like motif (I269V and/or N273D), or ii) occurrence of PAs.

For branch A, from which all *HSS*-like sequences of all PA-producing species descend, no sign of positive selection could be detected (Fig. 3, 4c, S3, Table 1). The likelihood of null and alternative model are equal (LRT=0). The substitution I269 to V269, resulting in formation of the full HSS motif, was reconstructed for Branch B in both trees, but this branch in both trees did not experience positive selection (LRT=0) (Figs. 3, 4d, S3; Table 1). In the full dataset topology, the substitution I269V was also reconstructed on the branches to the *Strophanthus boivinii, Isonema smeathmannii,* and the MRCA of Malouetieae *HSS*-like genes (Branch F) (Figs. 3, S3, Table 1). Although the calculated ω_2 was high for all three branches (ω_2=16.5, ω_2=197, ω_2=158, respectively, Table 1), positive selection was statistically significant only for Branch F after correction for multiple testing. In the reduced dataset, a parallel substitution of I269 to V269 in branch G, the MRCA of 10 *HSS*-like genes in Nerieae and Malouetieae, was reconstructed (Fig. 3a, b, 4c), but this branch did not experience positive selection (LRT=0) (Table 1).

We also tested branches leading to taxa known to produce PAs, which are the same branches in both the full and reduced tree topologies. The branch to *Anodendron HSS*-like

sequences (Branch D, Fig. 3b, 3d) experienced purifying selection in both trees (ω_1=0.133; ω_1=0.154, Table 1), as did the branch to *Amphineurion marginatum* (both topologies: ω_1=0.0515). The branch to the MRCA of these two lineages (*Amphineuron marginatum* and *Anodendron* spp., Branch E, Figs. 3b, 3d) also experienced strong purifying selection (both topologies ω_1=0.0). The branch to the MRCA of Echiteae (branch C, Figs. 3b, 3d) had a high ω2 (ω_2=18.7, ω_2=19.7, Table 1) on a small proportion of sites, but this was not statistically significant after correction for multiple testing (Table 1).

RELAX: Tests for shifts in selection intensity

We also used RELAX, which detects changes of selection strength (Wertheim, et al. 2015), to test our hypothesis that evolution of a more DHS-like motif from a more HSS-like motif in *HSS*-like genes is evidence of loss of PA biosynthesis, accompanied by relaxation of selection on HSS function and a shift toward neutral evolution (as predicted by the single origin, defense de-escalation model, Figs. 3a, c). Test clades where a more DHS-like motif (IXXXN or IXXXD) evolved from a more HSS-like motif (IXXXD or VXXXD) were compared to reference clades which retained the more HSS-like motif. None of the predictions derived from the defense de-escalation hypothesis were supported. In testing IXXXD branches (Clade K) against VXXXD branches in Clade G, which only occurred in the reduced dataset topology (Figs. 3a, 4c), the RELAX null model (logL=-34093.5, AIC_C=69335.7) was a better fit than the alternative model (LogL=-34093.5, AIC_C= 69337.6), supporting no shift in selection intensity (Table S11.1). Further, the alternative model indicated intensification of positive selection in Clade K (k=1.17) rather than relaxation, but it was not significantly better than the null model (k=1) (LR=0.14, p=0.711) (Table 2). In testing IXXXN branches (Clade I, Clade J, *Gongronemopsis truncata HSS*-like 2) against reference IXXXD branches in Clade L (Figs. 3a, c), the results differed

between tree topologies, and replicate runs for both tree topologies had variable results that had issues with convergence (Tables 2, S11.2) and warnings that results were untrustworthy, rendering all statistical inferences from these tests suspect. Similarly, in testing all IXXXN *HSS*-like branches against all IXXXD and VXXXD *HSS*-like branches (Clade A, Figs. 3a, c), results were not consistent between tree topologies (Tables 2, S11.3). This incongruence in results makes interpretation of selection intensity on IXXXN vs IXXXD/VXXXD branches impossible.

Lastly, we hypothesized that the APSA *HSS*-like genes should be under relaxed selection relative to *DHS*-like genes because DHS function is essential in primary metabolism while the rare occurrence of PAs predicts that *HSS*-like genes experienced frequent loss of optimized HSS function under the defense de-escalation hypothesis (Fig. 3a, c). The alternative model for both tree topologies was significantly better than the null model, supporting an intensification of positive selection on the *HSS*-like clade (Clade A ω_3=92.32. Clade H ω_3=32.05; Table S11.4). The intensification of selection was significant in the full dataset topology (k=1.31, LR=4.19, p=0.041) but not in the reduced dataset (LR=2.81, p=0.094). This can be interpreted as an overall stronger signal for positive selection for optimized HSS function than relaxation of selection on HSS function when averaged across all branches of the *HSS* subtree.

There is a pattern of relaxation of selection present in ancestral *DHS*-like and *HSS*-like branches, as illustrated by the optimal k-parameters inferred by the RELAX general descriptive model (Figs. 4, S3). After the core APSA *DHS/HSS* duplication event, there was a relaxation of purifying selection on the ancestral *DHS*-like Branch H (full dataset topology: k=0.09; reduced dataset: k=0.12) and the ancestral *HSS*-like Branch A (full dataset: k=0.01; reduced dataset: k=0.01) (Figs. 4b, S3; Tables S9, S10). The succeeding three branches in the backbone of the core *DHS*-like clade then experienced intensification of selection (full dataset: k=2.71-2.74;

reduced dataset: k=3.34-3.40), while most of the backbone of the core *HSS*-like clade continued to experience predominantly relaxation of selection (full dataset: k=0.04-0.80; reduced dataset: k= 0.08-1.94) (Figs. 4b, S3; Tables S9, S10).

MEME: Identification of candidate functional sites

We hypothesized that individual sites experienced positive selection on branches where optimized HSS function may have evolved. We tested branches (Figs. 3, 4, S3) selected based on motif and phenotype for evidence of site-specific positive selection using the branch-site model implemented in MEME (Murrell, et al. 2012). For the full dataset tree topology (Figs. 3c, d, S3) six amino acid sites showed evidence of significant episodic positive selection on these branches: sites 19 (LRT=4.7, p=0.04), 41 (LRT=5.04, 0.04), 51 (LRT=6.27, p=0.02), 75 (LRT=7.22, p=0.01), 246 (LRT=5.24, p=0.03), and 275 (LRT=9.77, p=0.00), and one site was trending towards significance: site 200 (LRT=4.22, p=0.06) (Table 3). For the reduced dataset tree (Figs. 3a, b, 4), only one amino acid site showed evidence of significant episodic diversifying selection on the test branches: 269 (LRT= 5.36, p=0.03), and one site is trending towards significance: site 19 (LRT=4.29, p=0.05) (Table 3).

Position 269, importantly, is the first amino acid in the focal motif (I/VXXXN/D). Of the remaining seven sites, two may be important for optimization of HSS function and/or loss of eIF5A activation based on substitution patterns. Position 75 is a synapomorphy for the *HSS*-like clade (D75A substitution on Branch A) in both trees (Figs. 4c, S3). Position 246, which is adjacent to an active site position in human DHS (Wator, et al. 2020), is 100% conserved in *DHS*-like sequences, including both Wrightieae paralogs. However, L246I substitutions occur 7 times in the *HSS*-like clade: on the branches leading to *Isonema smeathmannii* (Nerieae), *Galactophora schomburgkiana* (Malouetieae), the MRCA of Echiteae (Branch C), the MRCA of

Amphineurion+*Eucorymbia* (Apocyneae), the MRCA of Asclepiadeae (Branch I), the MRCA of Periplocoideae, and the MRCA of the *Mandevilla HSS*-like clade (Figs. 3, 4, S3). This same substitution, L246I, is present in all other functionally characterized angiosperm HSS except *Eupatorium cannabinum, Ipomoea alba, I. hederifolia, and I. neei* (Table S2).

None of the 8 amino acid positions indicated by MEME as experiencing positive selection, nor the second position (273) in the amino acid motif, were associated with active site positions in human DHS (Wator, et al. 2020) (Tables 3, S2). Position 19 aligns with amino acids within the human DHS ball-and-chain motif (Wator, et al. 2020). This position, as well as positions 269 and 273 in the DHS/HSS motif, all align with amino acids involved in protein-protein interactions in the quaternary structure of the human DHS homotetramer (Wator, et al. 2020). Positions 246, 269, 273, and 275 all align with amino acids associated with the tunnel entrance area on the human DHS dimer (Wator, et al. 2020).

Function of HSS VXXXD motif

To further study the functional relevance of the VXXXD motif in Apocynaceae HSS, we constructed single (IXXXD, VXXXN) and double (IXXXN) mutants of the HSS from *Parsonsia alboflavescens* [a PA producing species (Barny, et al. 2021)] and tested their activity with putrescine to produce homospermidine (HSS function, Fig. 1B) and the eIF5A precursor protein to produce activated eIF5A with a lysine aminobutylated to deoxyhypusine (DHS function, Fig. 1A) in *in vitro* assays. All four enzymes were successfully expressed and purified (Fig. S5). SEC-UV confirmed that the expressed enzymes assemble to a tetramer, which is the biologically active unit of HSS (Umland, et al. 2004; Wator, et al. 2020).

To compare HSS activity among the enzymes, the specific activity for homospermidine production was calculated (Table 4). The activity of the IXXXD single mutant is comparable to

that of wild type VXXXD HSS (mean 1145±34.35 pkat/mg, 1317±210.72 pkat/mg, respectively). Both the VXXXN single (mean 471±4.71 pkat/mg) and the IXXXN double (mean 526±36.82 pkat/mg) mutants show a decreased specific activity compared to the wild-type enzyme. 1,3-diaminopropane (dap) was produced in roughly equimolar amounts with homospermidine, indicating that the transfer of the aminobutyl moiety to putrescine was the major reaction catalyzed by the enzymes. HSS activity is high in all four enzymes when compared to *P. alboflavescens* DHS, 274 pkat/mg (Livshultz, et al. 2018). The two-fold higher activity measured for the IXXXD and VXXXD enzymes compared to the VXXXN and IXXXN enzymes may be due to loss of catalytic activity caused by the D273N substitution (suggesting functional importance of this substitution for HSS optimization) or to a greater ratio of inactive to active enzyme in the latter two enzyme preparations (suggesting an experimental artifact).

To assess DHS activity, dap formation was quantified in the presence of eIF5A precursor protein and compared to a negative control assay without it (Table 4). Wildtype and IXXXD mutants produced similar amounts of dap and canalvamine (Can) in both assays, indicating that both enzymes do not aminobutylate the eIF5A precursor protein (Table 4, Fig. S6) but instead use spermidine (spd) as aminobutyl acceptor. Of note, while Can increased constantly over the time, dap was produced in a burst at the beginning of the reaction, which is very pronounced for the IXXXD mutant (Fig. S6). This burst most likely results from spd cleavage to dap and delta pyrrolinium, without transfer of the aminobutyl moiety (Fig. S6). The VXXXN mutant in the assay with the eIF5A precursor produced slightly less dap (mean 41±3.29 pkat/mg) than when eIF5A precursor was excluded (mean 50±2.5 pkat/mg). In contrast, the double mutant IXXXN produced more dap and can when eIF5A was included. Of note, deoxyhypusine, evidence of DHS function, could only be detected when VXXXN and IXXXN but not VXXXD or IXXXD

enzymes were incubated with the eIF5A precursor protein. Thus, the N273D substitution might be important for interaction of the enzyme with eIF5A. It also affects the rate of spermidine cleavage, which is highest in the IXXXD mutant.

Discussion

Detecting adaptive evolution of phenotypic characters is a major challenge in evolutionary studies above the population level because natural selection acts on variation within populations. Coddington (1988) proposed that evidence of superior function in the apomorphic compared to the plesiomorphic state of a character is the best evidence we can adduce to the problem. Other researchers have suggested that evidence of positive selection on genetic loci that underlie a phenotypic trait can also test the hypothesis that a phenotypic trait evolved under positive selection (Rausher and Delph 2015).

Livshultz, et al. (2018) applied Coddington's criteria to test the defense de-escalation hypothesis of pyrrolizidine alkaloid evolution in Apocynaceae: that PAs evolved as an anti-herbivore adaptation once early in the diversification of the family, followed by multiple loses in lineages highly exploited by PA-philous insects. They showed 1) that the duplication that gave rise to the first gene of PA biosynthesis, homospermidine synthase, arose early in the diversification of one lineage, the APSA clade; 2) that a characteristic HSS motif, VXXXD, evolved from the ancestral IXXXN motif in this ancestral paralog; and 3) that this motif reverted to IXXXN or IXXXD and/or the *HSS*-like gene was pseudogenized in species not known for PA biosynthesis. However, Livshultz, et al. (2018) lacked direct evidence of the functional importance of the HSS VXXXD motif.

In the present study, we revisit the Livshultz, et al. (2018) analysis with an expanded dataset including new sequences of Apocynaceae *HSS*- and *DHS*-like genes, new phenotypic

data on PA occurrence (Barny, et al. 2021), analyses of selection patterns on these genes, and new data on the effect of motif on enzyme function. We show that the reconstruction of the evolution of *HSS*-like genes by Livshultz, et al. (2018) was fundamentally correct, but patterns of selection on *HSS*-like genes do not, overall, support the defense de-escalation hypothesis (Fig. 3a, c). We conclude that the evolutionary history of PAs cannot be inferred from evolution of *HSS*-like genes alone because of substantial uncertainty in the genotype to phenotype map (Weiss and Fullerton 2000; Weiss 2010). We identify future research that can reduce this uncertainty.

An HSS-*like gene with a highly functional motif evolved early in the diversification of the APSA clade, followed by multiple independent losses*

An HSS-like enzyme with an IXXXD motif evolved in the APSA clade after the divergence of Wrightieae. While this is not the VXXXD motif as reconstructed by Livshultz, et al. (2018), all motifs have high HSS activity in the function assays (Table 4). Therefore, the MRCA of all PA-producing Apocynaceae could have produced abundant homospermidine. Additionally, the ancestral DHS activity with the eIF5A precursor (Fig. 1A) would have been lost in the IXXXD ancestral HSS, as evidenced by the restoration of DHS activity in IXXXN and VXXXN HSS mutants (Table 4). Evolution of a putatively highly functional HSS-like enzyme early in the diversification of the APSA clade, was followed by multiple losses of function, as inferred by evidence of pseudogenization of all *HSS*-like genes in 6 independent lineages (*Apocynum*, *Asclepias syriaca, Nerium oleander, Pachypodium barronii, Strophanthus preussii, Zygostelma benthamii*) (Fig. 2). However, the implication of evolutionary loss of the IXXXD or VXXXD motif for homospermidine production is not as apparent since the IXXXN double mutant still has high catalytic activity with putrescine (Table 4). The IXXXN double mutant HSS

of *Parsonsia alboflavescens* is also able to perform DHS-like activity, aminobutylation of eIF5A, unlike the VXXXD wild-type and IXXXD single mutants, and this DHS-like motif evolved at least twice: in the MRCA of the Asclepiadeae tribe (Branch I) and in one of the two Marsdenieae *HSS*-like paralogs (Branch J) (Figs. 3, 4e, S3.

The absence of positive selection on the *HSS*-like gene in the MRCA of PA-producing Apocynaceae (Branch A) (Figs. 4c, S3; Table 1) does not disprove a single origin of a highly functional HSS and PA biosynthesis. *De novo* evolution of high HSS function could occur under relaxed selection (Weng, et al. 2012; Rey, et al. 2019). After gene duplication, the duplication-degeneration-complementation model predicts that both paralogs will experience relaxation of selection, allowing potentially adaptive mutations to accumulate in one of the paralogs (Force, et al. 1999; Ober and Kaltenegger 2009; Panchy, et al. 2016). The MRCA of core APSA *HSS*-like (Branch A) and *DHS*-like (Branch H) paralogs both experienced relaxed selection (Figs. 4b, c, S3; Tables S9, S10), and a functionally important amino acid substitution (N273D) which abolished DHS activity and may have increased HSS activity occurred on this ancestral *HSS*-like branch (Branch A) (Table 4), consistent with the DDC model. Therefore, if the entire PA biosynthetic pathway evolved at or around the same time as this ancestral IXXXD-motif HSS, a single origin of PAs and defense de-escalation is supported.

Selection tests weakly favor convergent evolution of optimized HSS

There is weak evidence from selection tests to support convergent evolution of optimized HSS (and by inference, PA biosynthesis) in different Apocynaceae lineages. The signal of intensifying selection on the entire *HSS*-like clade when compared to the *DHS*-like clade along with a pattern of intensification of selection on multiple branches in the *HSS*-like clade is more consistent with multiple origins of functionally optimized HSS. There is a trend toward positive

selection, though not statistically significant when corrected for multiple testing, on Branch C, MRCA of *HSS*-like sequences from PA-producing Echiteae, in both trees (Figs. 3b, 3d, 4d, S3). Though not formally tested, the MRCA of *Alafia* (Branch K) had a very high ω on a minority of amino acid sites in both the reduced dataset tree (ω_1=0.363, 99.48% sites; ω_2=62.5, 0.52% sites) and the full dataset topology (ω_1=0.353, 99.46% sites; ω_2=61.4, 0.54% sites) (Tables S9, S10). There is statistically significant positive selection on the ancestral branch (Branch F) of *HSS*-like sequences from Malouetieae, a tribe that includes a species reported as PA-producing (Arseculeratne, et al. 1981), but there is still only low to moderate confidence in presence of PAs among the species tested by Barny, et al. (2021). It remains unclear, however, when (or if) significant positive selection to optimize HSS function occurred in the ancestors of the known PA-producing genera in Apocyneae: *Anodendron* and *Amphineurion*. The MRCA of *Amphineurion*+*Eucorymbia* does have high ω2 in the full dataset topology (ω_1=0.272, 99.32% sites; ω_2=543, 0.68% sites) and the reduced dataset tree (ω_1=0.273, 99.34% sites; ω_2=670, 0.66% sites) (Tables S9, S10). Similar to Malouetieae, *Eucorymbia* may produce PAs but the current evidence is not conclusive (Barny, et al. 2021).

Positive selection on HSS may not necessarily indicate *de novo* evolution of PA biosynthesis. It could instead be a signal of HSS optimization under selection from herbivores for increased PA defense after PA biosynthesis had already evolved. The latter scenario has been proposed for the pattern of selection on a defensive specialized metabolic enzyme, threonine deaminase (TD2) (Rausher and Huang 2016). Rausher and Huang (2016) suggest that a 25-30 my span of episodic positive selection on TD2 was a result of fluctuating herbivory: high abundance of herbivores selected for adaptive evolution of TD2 and a decrease in herbivore abundance relaxed selection on TD2.

Furthermore, positive selection on HSS may stem from selection against paralog interference with DHS (Kaltenegger and Ober 2015) as well as selection for optimized catalytic function. Despite evidence for significant positive selection for the I269V substitution in Apocynaceae (Table 3) and Convolvulaceae HSS (Kaltenegger, et al. 2013), experimental amino acid substitution at position 269 in the wild-type *Parsonsia alboflavescens* HSS (V269I) did not significantly change catalytic activity (Table 4). This position, as well as the other motif position (N273D), is homologous to amino acids involved with hydrogen bonding between the dimers of human DHS (Wator, et al. 2020). Amino acid changes at these positions could possibly prevent the heterodimerization of DHS and HSS paralogs, enabling the two paralogs to perform their primary functions optimally (Ober and Kaltenegger 2009).

Limitations of selection tests: potential for incorrect classification of branches

The validity of our selection tests depends on the correct identification of branches where HSS function either increased or decreased. We used a single motif as a proxy for HSS function, which did not allow us to examine the contributions of other potentially functionally important sites (Table 3) and we have no knowledge of expression levels. We also used presence of PAs to infer highly functional *HSS* genes. However, the occurrence of PAs is still under-determined in Apocynaceae (Barny, et al. 2021), i.e. some species in our analysis are likely to be misidentified as lacking PAs. Better understanding of structure-function relationships in HSS and PA distribution in Apocynaceae are necessary for future refinement of these analyses. Furthermore, new models may better capture the evolutionary process by simultaneous reconstruction of selection on genotype and phenotype, rather than on genotype alone. For example, a shortcoming of RELAX is the *a priori* designation of foreground and background branches, as is the assumption that trait changes occur at nodes (i.e. speciation events) rather than elsewhere on a

branch (Wertheim, et al. 2015; Halabi, et al. 2020). This is being improved upon, for example by TraitRELAX, which allows trait transitions to occur multiple times across a branch and uses k (selection intensity parameter) based on these reconstructions to designate foreground and background branches (Halabi, et al. 2020). Halabi, et al. (2020) found that TraitRELAX performed better than RELAX. While it can presently only handle phenotypic binary traits, the understanding of the PA phenotype in Apocynaceae is far too fragmentary for use in ancestral phenotype reconstruction.

Genotype-phenotype map of HSS motif and PA presence indicates partial disconnect between motif-predicted function and PA phenotype

Despite the comparable functional level of an IXXXD HSS mutant and the VXXXD wild-type enzyme of *Parsonsia alboflavescens* (Table 4), no species with an IXXXD HSS-like motif has been shown to produce PAs. Additionally, many taxa with an intact *HSS*-like gene encoding a VXXXD motif do not have evidence of PAs (Barny, et al. 2021, 2022). Of the 17 Apocyneae (VXXXD) genera tested (of 21 genera total), PAs have been found in only two, *Anodendron* and *Amphineurion* (Sasaki and Hirata 1970; Colegate, et al. 2016). Within the MOE polytomy, PAs have only been found in Echiteae (Barny, et al. 2021). Presence of PAs in Malouetieae (VXXXD) is currently uncertain and requires further testing (Barny, et al. 2021). There is therefore a disconnect between our definition of a highly functional HSS (IXXXD or VXXXD motif) and PA phenotype. Setting aside error in phenotypic classification (Barny, et al. 2021), there are many levels at which this disconnect may occur.

Substitutions at the two focal amino acid sites (positions 269 and 273) may be insufficient to predict HSS function. We identified additional candidate functional sites homologous to functional sites in human DHS, based on evidence of positive selection (Table 3).

Position 19 in Apocynaceae DHS/HSS-like enzymes is predicted to be part of the ball-and-chain motif (Liao, et al. 1998; Umland, et al. 2004; Wator, et al. 2020). Presence of this motif is necessary for DHS to function, but this is the least conserved part of DHS and amino acid changes could accumulate at a higher rate without altering the function of the ball-and-chain motif (Umland, et al. 2004; Wator, et al. 2020). Five of the amino acids of particular interest (positions 74, 75, 246, 269, 273, identified via selection and/or functional assays) are linked to functional areas in human DHS. Positions 246, 269, and 273 are predicted to be near the enzyme surface at the active site tunnel entrance, so binding substrates likely encounter these amino acids (Prakashrao, et al. 2022). Umland, et al. (2004) showed that the human DHS tunnel entrance had a less negative charge than the modelled *Senecio vulgaris* (Asteraceae) HSS, leading to the prediction that certain amino acid substitutions in this area could affect binding affinity. Relative to its DHS, *S. vulgaris* HSS has the same substitutions at the active site tunnel entrance (L246I, I269V, N273D) as Apocynaceae HSS. It is therefore likely that the same conclusions Umland et al. (2004) made about *S. vulgaris* HSS apply to Apocynaceae. Considering that the putative active sites of all intact Apocynaceae HSS-like amino acid sequences are conserved, it theoretically follows that functional changes could result from external amino acids that deflect bulky eIF5A (i.e. DHS-like function), but allow the entrance of much smaller polyamines (spermidine and putrescine; i.e. HSS-like function) (Umland, et al. 2004; Ober and Kaltenegger 2009; Wator, et al. 2020; Prakashrao, et al. 2022). After experimental validation of the effect of these substitutions on function, a combination of these sites may be more effective at identifying functionally optimized HSS.

Incongruence between *HSS* genotype and PA phenotype could result from mutations in promoter regions that reduce expression of *HSS*-like genes with intact ORFs and VXXXD

motifs. In neotropical bats (Noctilionoidea), intact S-opsin genes were sequenced from multiple lineages that had lost S-opsin function and the ability to see in the UV spectrum (Sadier, et al. 2018). Sadier et al. (2018) found evidence of multiple independent losses of transcription or translation of S-opsin genes with intact ORFs. The same may be obtained for *HSS*-like genes of PA-free Apocynaceae. Testing this possibility requires studies of gene expression and protein accumulation.

Similarly, incongruence can result from enzymes downstream of HSS in the PA biosynthetic pathway. *HSS* is the first gene of the PA biosynthetic pathway, but pseudogenization or loss of transcription or translation could occur in any gene of the pathway. Based on simulations, Wright and Rausher (2010) proposed that in a linear specialized metabolic pathway, the first enzyme is more likely to retain adaptive substitutions (e.g. N273D). Downstream enzymes are more likely than first enzymes to retain slightly deleterious amino acid changes (Wright and Rausher 2010), thus making it less likely for evidence of PA biosynthesis loss to be detectable from looking only at HSS. The PA biosynthetic pathway may be a more complex branching pathway. But, each sub-branch within a branching pathway acts as a linear metabolic pathway: the first enzyme after branching is more likely to retain adaptive substitutions and less likely to retain deleterious ones than downstream enzymes (Rausher 2013). Therefore, if HSS is the first enzyme of a branch in a complex pathway, the breakdown of the PA biosynthetic pathway is still more likely to occur in a downstream enzyme. The predicted importance of enzyme order to the evolution of pathway function can be tested when downstream genes in the PA biosynthetic pathway are identified and sequenced across Apocynaceae.

The apparent disconnect between genotype and phenotype could be because HSS has other functions besides producing PA precursors, e.g. homospermidine may be a precursor to specialized metabolites aside from PAs (Bienz, et al. 2002).The optimization of HSS function could have predated the evolution of PA biosynthesis, resulting in Apocynaceae lineages with highly functional HSS enzymes but no PAs. In *Bicyclus anynana* (African squinting bush brown butterfly), multiple duplication events resulted in the *engrailed* (*en*) gene family, expressed in the eyespots of these butterflies; however, these duplication events occurred 60 my before the evolution of eyespots (Banerjee, et al. 2020). Banerjee, et al. (2020) concluded that since the evolution of e*n*-gene family did not coincide with eyespot evolution; the duplicates were therefore retained for other reasons. The retention of HSS-like enzymes in non-PA-producing Apocynaceae could alternatively be related to DHS function (i.e. activation of eIF5A), rather than HSS function, via re-recruitment as a DHS. As in Apocynaceae, only a minority of species in Convolvulaceae with *HSS*-like genes produce PAs (Prakashrao, et al. 2022). Also similar to the results here, the IXXXN HSS *in Ipomoea alba* (Convolvulaceae), a non-PA producing species within a genus that contains PA-producing species, is able to produce deoxyhypusine, unlike Convolvulaceae species with either IXXXD or VXXXD HSS (Prakashrao, et al. 2022). Further afield, Afanador, et al. (2018) discovered a complex evolutionary history of *Trypanosoma brucei* (Euglenozoa) DHS. The DHS in *T. brucei* is a heterotetramer comprised of a catalytically impaired DHS paralog (IXXXN), which contains the spermidine binding positions in the active site, and a catalytically dead (VXXXR) paralog, which contains the NAD+ binding sites (Afanador, et al. 2018). A heterotetramer comprised of these two paralogs is over 1,000-fold more catalytically active than a homotetramer of the catalytically impaired DHS (Afanador, et al. 2018). The possibility that *HSS* orthologs, particularly those that have re-evolved the

IXXXN motif, in non-PA producing Apocynaceae function as DHS loci can be tested using DHS activity assays on these enzymes.

Alternatively, the evolution of an optimal or highly functional HSS may post-date evolution of the PA biosynthetic pathway. DHS is a promiscuous enzyme that has at least two functions, the secondary of which produces homospermidine (Fig. 1). Trace amounts of homospermidine, presumably produced by DHS, are ubiquitous in plants (Ober and Hartmann 1999b; Ober, Gibas, et al. 2003). The recruitment of a promiscuous enzyme into a specialized metabolite biosynthetic pathway would occur when the latent product becomes relevant, and this theoretically could occur before, simultaneous to, or after gene duplication (Noda-Garcia and Tawfik 2020). Specialist enzymes are often reconstructed as having evolved from generalist, promiscuous enzymes (reviewed by Gumulya and Gillam 2017). Therefore, DHS could have been recruited to a specialized metabolic pathway and persisted in its substrate promiscuity until specialization of HSS, i.e. inability to bind to eIF5A and increased production of homospermidine for high PA biosynthesis, was selected for (Kreis and Munkert 2019; Noda-Garcia and Tawfik 2020).

Future work: Reducing uncertainty in the genotype-phenotype map to clarify evolution of PA biosynthesis in Apocynaceae

Despite a detailed reconstruction of the *DHS/HSS* gene tree, there are clearly gaps in data that must be filled to more definitively test hypotheses concerning the evolution of PA biosynthesis in Apocynaceae. First, a better understanding of the distribution of phenotypes and HSS-like enzyme functions is necessary, including functional analyses of reconstructed ancestral enzymes and expression studies. Analyses of synteny can improve inferences about gene orthology that remain ambiguous because of lack of support in the gene tree topology, e.g.

orthology of the Wrightieae *DHS*-like 2 locus. Secondly, additional candidate genes of the PA biosynthetic pathway need to be identified using coexpression networks (Wisecaver, et al. 2017). The second enzyme, a member of the copper-containing amine oxidase (CuAO) family, in the PA biosynthetic pathway has recently been identified in *Heliotropium indicum* (Boraginaceae) (Zakaria, et al. 2022). The identity of this gene and other candidate genes can be predicted by studying expression in relation to presence of PAs (Donoso, et al. 2021). *Echites umbellatus*, which has variable PA production within populations and within individual plasticity (Barny, et al. 2021) would be a good species for this study. Genome mining (Franke, et al. 2019) combined with functional assays can then quickly identify functional orthologs across related species. When additional genes in the PA biosynthetic pathway are identified in Apocynaceae, evolution of the entire pathway can be reconstructed (Piatkowski, et al. 2020). Assembly of an entire biosynthetic pathway provides much stronger evidence for placing the evolution of a metabolite than evolution of a single enzyme of the pathway. Pathway evolution then needs to be reconstructed on a species tree. The most recently published Apocynaceae phylogeny (Fig. 2), based on 353 nuclear loci and sparse taxonomic sampling, was not able to resolve the APSA clade rapid radiation (Antonelli, et al. 2021) but additional gene and taxon sampling may do so.

Materials and Methods

Taxon sampling for DHS/HSS sequencing

DNA samples were collected for 170 samples of 159 species in Apocynaceae (Table S1), including samples from known PA-producing species (*Amphineurion marginatum, Anodendron affine, Echites umbellatus, Laubertia boissieri, Macropharynx peltata, Parsonsia alboflavescens, P. eucalyptophylla, Prestonia portobellensis, Rhodocalyx rotundifolius, Temnadenia odorifera*)

(del Castillo, et al. 1970; Sasaki and Hirata 1970; Hartmann, et al. 2001; Pasteels, et al. 2001; Burzynski, et al. 2015; Colegate, et al. 2016; Tasca, et al. 2018; Barny, et al. 2021). Also represented in this sampling is an outgroup species, *Gelsemium sempervirens* (Gelsemiaceae) (Antonelli, et al. 2021). Previously published Apocynaceae *DHS* and *HSS* sequences from 25 species were also included in analyses (Livshultz, et al. 2018).

Calotropis genome query

Dhs- and *HSS*-like genes were extracted from the *Calotropis gigantea* genome (Hoopes, et al. 2018) via tblastx searches with exon sequences of *Parsonsia alboflavescens DHS* and *HSS* (MG817648.1, MG817649.1) in Geneious Prime v.2020.0.3 (https://www.geneious.com).

DNA extraction, library preparation, sequencing

DNA extraction, library preparation, and paired end sequencing follow protocol described in Fishbein, et al. (2018), with the addition of enrichment for *DHS* and *HSS* by Arbor Biosciences using a recently published Apocynaceae nuclear gene probe set (Straub, et al. 2020). Contaminated libraries (i.e. libraries containing DNA from more than one sample) were identified by mapping a sample's reads to its own assembled plastome. Samples with evidence of two divergent plastomes were considered contaminated and excluded (unpublished data).

Contig assembly

Raw paired end sequence reads were trimmed using Trimmomatic default settings (slidingwindow:10:20, minlen:40) (Bolger, et al. 2014). Using the first stage of the MyBaits pipeline [BLASTN (Cameron and Williams 2007), SPAdes (Bankevich, et al. 2012)] with default options (SPAdes: k 21,33,55,77, cov-cutoff off, phred-offset 33) (Moore, et al. 2018), trimmed sequence reads for each sample were then binned as *DHS*/*HSS* by BLASTN using reference transcriptomic *DHS*-like sequences (Table S1) and binned reads were assembled using

SPAdes. Libraries that did not have an assembled contig with > 20x coverage were removed from analysis (Bentley, et al. 2008; Dohm, et al. 2008; Harismendy, et al. 2009; Whittall, et al. 2010; Straub, et al. 2012). Retained SPAdes contigs were extended and fused by afin, an assembly finishing program (-s 50, -l 100) (McKain and Wilson 2017).

Exon annotation

Retained afin contigs were annotated for exonic regions in Geneious Prime v.2020.0.3 (https://www.geneious.com) using *DHS* and *HSS* exons from *Parsonsia alboflavescens* (GenBank accessions: MG817648.1, MG817649.1) (50% identical threshold). Annotated contigs (exon and intron) were aligned within a sample to identify and manually annotate partial or divergent exons that did not meet identity threshold.

Alignment construction

All alignments were constructed initially as nucleotide alignments using MAFFT v7.450 (gap penalty=3) (Katoh, et al. 2002; Katoh and Standley 2013) in Geneious Prime v.2020.0.3 (https://www.geneious.com).

ML tree construction

All maximum likelihood trees were constructed using the same parameters, with rapid bootstrapping and partitioned by codon position in RAxML-HPC v.8 on XSEDE (v8.2.12) (GTR+GAMMA, 1000 bootstrap replicates) (Stamatakis 2014) in CIPRES Science Gateway (v3.3) (Miller, et al. 2010).

Gene assembly from contigs

Annotated exons were extracted and aligned with an existing Apocynaceae *DHS/HSS*-like sequence alignment (Livshultz, et al. 2018) and *Calotropis gigantea* sequences (Hoopes, et al. 2018). A maximum likelihood tree was built from this “initial alignment” (Fig. S1, Tables S1,

S2). Contigs orthologous to *Parsonsia alboflavescens HSS* were considered *HSS*-like. Contigs outside of this clade were considered *DHS*-like. Contigs were concatenated using the following algorithm. After excluding contigs with non-terminal stop codons, i.e. potential pseudogenes, a strict consensus was calculated from overlapping partial contigs that were in a clade with other *DHS*-like or *HSS*-like sequences from the same tribe in the initial alignment gene tree (Fig. S1) from the same library. Amino acid similarity in the region of overlap of the combined contigs ranged from 89.6-100% (Table S1). Then non-overlapping contigs were concatenated using the same grouping criteria. Polyphyletic contig pairs were grouped if there were no sequences from closely related taxa for them to cluster with, and there was no evidence of contamination in the source library.

Validation of gene assemblies

Sequence assemblies for samples that had Sanger sequences of the same locus from Livshultz, et al. (2018) were validated via pairwise alignment and calculation of divergence. Here, and otherwise, divergence/pairwise sequence identity was calculated in Geneious using nucleotide alignments; this excludes missing data.

Construction of matrices

Sanger sequences >95% identical to new sequence from the same sample were removed from the alignment. The "Full Dataset" consisted of potential pseudogene sequences, full (minimum exons 2-6) *DHS*-like and *HSS*-like sequences, short *DHS*-like and *HSS*-like contigs (from samples in which an entire *DHS*-like or *HSS*-like gene had also been assembled by SPAdes/afin), and non-redundant Sanger sequences from Livshultz, et al. (2018) (Tables S1, S2). This full dataset alignment was trimmed to produce a "Reduced Dataset" alignment used for ancestral state reconstruction and selection analyses. Potential pseudogenes and sequences that

did not span at minimum exons 2-6 were removed from this dataset and remaining sequences were realigned. Thirty-seven (37) bp were trimmed from the 5' end and 67 base pairs were trimmed from the 3' end of this reduced dataset alignment, as this area was highly variable.

Gene tree construction

Three gene trees were produced: a full dataset tree was constructed that contains all contigs (candidate pseudogenes, short contigs, and full/concatenated/consensus contigs) (Fig. S2), and a reduced dataset tree (only containing contigs with at least exons 2-6 and no non-terminal stop codons) (Fig. 4). Lastly, a third tree was constructed using the reduced dataset alignment 100% constrained to the full dataset tree topology, referred to as the "full dataset topology" (Fig. S3).

Tests for recombination

Marsdenieae *HSS*-like paralogs were investigated further using GARD (Kosakovsky Pond, et al. 2006). GARD searches an alignment for a maximum number of breakpoints, builds phylogenies for every non-recombinant contig, and assesses those phylogenies using Akaike Information Criterion (AIC). Potentially recombinant contigs were split at potential recombination breakpoints indicated by GARD and a Marsdenieae *HSS*-like gene tree (outgroup: *Tassadia propinqua HSS*-like gene) was rebuilt using maximum likelihood tree construction criteria described above.

Shimodaira-Hasegawa tests

To test alternate topologies of Apocynaceae *DHS-/HSS*-like gene trees and test the monophyly of IXXXN and IXXXD paralogs in Marsdenieae, Shimodaira-Hasegawa (SH) tests in RAxML-HPC2 on XSEDE (v8.2.12) (Stamatakis 2014) were performed. The SH test rejects or fails to reject a null hypothesis of equal support for two given topologies.

Ancestral sequence reconstruction

Ancestral sequences were reconstructed using codeml Model M0 (default except: model=0, NSsites=0, RateAncestor=1, cleandata=0) in PAML v4.9j (Yang 1997; Yang 2007). codeML integrates the Goldman and Yang model of amino acid substitution and assumes that selection pressure on an individual site is the same for every branch, produces joint likelihood reconstructions (all ancestral nodes reconstructed), and uses empirical Bayes procedure (Yang and Wang 1995) for sequence reconstruction. Additionally, codeml calculates a marginal reconstruction (single nodes reconstructed), which includes posterior probabilities for reconstructed amino acids.

Tests for selection

Two different models in the HyPhy suite were used to test for evidence of positive selection in *HSS*. aBSREL (adaptive Branch-Site Random Effects Likelihood) (Smith, et al. 2015) takes short branch length into account, fitting optimal ω (dN/dS ratios) distributions to each branch by assigning one of three ω rate categories to each branch and each site on that branch. This generates the optimal ω for each branch, which is displayed in Figs. 4 and S3. Positive selection is inferred on the *a priori* test branches by comparing, via likelihood ratio test (LRT), the optimized ω to a null, ω <1. MEME (Mixed Effects Model of Evolution) allows selection to vary both among branches and among sites (Murrell, et al. 2012). Each branch is assigned one of two ω rate classes at each amino acid site. A single α (synonymous substitution rate) is shared among all branches. β- (1st nonsynonymous substitution rate) is estimated for each site; β- is constrained to be less than or equal to α (i.e. evolving neutrally). β+ (2nd nonsynonymous substitution rate) is unconstrained in the full model. Likelihood ratio tests are used to compare the full model with a null model where β+ is constrained to be less than α.

RELAX (Wertheim, et al. 2015) was used to test for relaxation and intensification of positive and purifying selection on selected test branches compared to a set of designated reference branches. The RELAX null model assigns all sites into one of three rate classes (ω1=purifying, ω2=neutral, ω3=positive selection). The full (alternative) model introduces a selection intensity parameter, k, and raises ω^k on the test branches. The null model constrains k=1, which forces the same ω distribution in both the test and reference branch sets. If the likelihood ratio tests find the alternative model is significantly better fit, a k>1 is considered evidence of intensified selection and a k<1 relaxed selection in the test branches relative to the designated reference branches. Additionally, the RELAX general descriptive model was used to produce the k values used in Figs. 4 and S3, which specifically came from the RELAX runs comparing *hss*-like branches to *dhs*-like branches (Table S11.4). Rather than using the *a priori* test and reference branch sets, the general descriptive model fits the three ω rates to all branches, and an individual k for each branch (Wertheim, et al. 2015).

Comparison of human DHS

The reduced dataset (described above) was aligned with the angiosperm *DHS/HSS* alignment from Livshultz, et al. (2018) and human *DHS* (Genbank: P49366) in Geneious (protocol described above) to produce a "human and plant *DHS/HSS* alignment" (Table S2). Human DHS amino acid site functions are described in Wator, et al. (2020). The amino acid positions of DHS monomer interaction sites and general functional sites (i.e. active site tunnel entrance) were inferred from annotations on structures PB ID 6XXXM_A, 6XXXM_A1, 6XXXM_B, and 6XXXM_B1 using the NCBI Structure feature (Madej, et al. 2014) on iCN3D v.2.24.4 (Wang, et al. 2020). These sites were manually annotated on the human *DHS* sequences

in the alignment to enable comparisons between human and plant DHS/HSS amino acid positions.

Site-directed mutagenesis Parsonsia alboflavescens *HSS*

The open reading frame of *Parsonsia alboflavescens* HSS (*Pa*HSS), cloned in an expression vector (NovagenTM pET28a, Millipore Sigma, Billerica, MA, USA) with an artificial N-terminal hexahistidine (6xHis) tag extension, was used as template for site-directed mutagenesis guided by Liu and Naismith (2008). Primer pairs to introduce the single mutations V269 to I273 (numbering of the amino acids follows Kaltenegger, et al. 2013) and D273 to N273 as well as to double mutation V269D273 to I277N281 are given in Table S3. PCR amplifications were performed in 25 µl reaction volume with Phusion® High-Fidelity DNA Polymerase (Thermo-Scientific) according to the manufacturer's instructions; annealing temperature is given in Table S3. Twelve (12) amplification cycles were performed. The PCR products were treated with *Dpn*I (Thermo-Scientific) at 37°C for 1 hour, diluted with water (1:10), subsequently propagated in *Escherichia coli* TOP10 (Thermo-Scientific), and sent out for Sanger sequencing (MWG Eurofins Genomics) to identify successful mutants.

Heterologous expression, purification and activity assays of P. alboflavescens HSS and mutants

The complete ORF of the *Pa*HSS and the mutant variants were expressed in *Escherichia coli* BL21(DE3) and purified according to (Ober and Hartmann 1999a). Protein purification was monitored via SDS-PAGE analysis and protein quantities were estimated based on UV absorption at 280 nm and the specific extinction of the respective protein, calculated with the PROTPARAM web tool on ExPASy (Gasteiger, et al. 2005) and with the Bradford method (Bradford 1976).

For biochemical characterization, the purified proteins were concentrated and rebuffered to borate-based (50 mM borate-NaOH buffer, pH 9) assay buffer which included the additives DTT (1 mM) and EDTA (0.1 mM). The *in vitro* assays were performed according to (Kaltenegger, et al. 2021). In short, 5 – 40 µg purified recombinant protein were incubated with putrescine and spermidine (400 µM each), in the presence of NAD (2 mM), in borate -based assay buffer to determine the enzyme's ability to produce homospermidine. Product formation was quantified via derivatizing the reaction mixture with 9-fluorenylmethyl chloroformate (FMOC, Sigma) and subsequent analyses by HPLC coupled with UV detection. To detect the enzymes' ability to utilize the eIF5A, assays were hydrolyzed as described in Kaltenegger, et al. (2021) derivatized with FMOC and analysed by HPLC coupled with FLD to quantify deoxyhypusine, 1,3-diaminopropane and canavalmine.

Acknowledgements

The authors thank Mark Fishbein (Oklahoma State University), André Simões (Universidade Estadual de Campinas), and Mary Endress (University of Zurich) for kindly sharing plant leaf tissue. We thank Hobart and William Smith Colleges undergraduates Madison Cullinan, Meredith Steinfeldt, Abbey Foote, Katarina Kostović, and Charmaine Chung for laboratory assistance. This work was supported by the National Science Foundation, United States, grant DEB-1655663 and DEB-1655223 to TL and SS and a Drexel University Clinical and Translational Research Institute (CTRI) seed grant to T.L.

References cited

Afanador GA, Tomchick DR, Phillips MA. 2018. Trypanosomatid Deoxyhypusine Synthase Activity Is Dependent on Shared Active-Site Complementation between Pseudoenzyme Paralogs. Structure 26:1499-1512 e1495.
Agrawal AA, Fishbein M, Halitschke R, Hastings AP, Rabosky DL, Rasmann S. 2009. Evidence for adaptive radiation from a phylogenetic study of plant defenses. PNAS 106:18067.
Ali JG, Agrawal AA. 2012. Specialist versus generalist insect herbivores and plant defense. Trends Plant Sci 17:293-302.
Alverson AJ. 2007. Strong purifying selection in the silicon transporters of marine and freshwater diatoms. Limnol. Oceanogr. 42:1420-1429.
Anke S, Niemüller D, Moll S, Hänsch R, Ober D. 2004. Polyphyletic origin of pyrrolizidine alkaloids within the Asteraceae. Evidence from differential tissue expression of homospermidine synthase. Plant Physiol 136:4037-4047.
Antonelli A, Clarkson JJ, Kainulainen K, Maurin O, Brewer GE, Davis AP, Epitawalage N, Goyder DJ, Livshultz T, Persson C, et al. 2021. Settling a family feud: a high-level phylogenomic framework for the Gentianales based on 353 nuclear genes and partial plastomes. American Journal of Botany.
Arseculeratne SN, Gunatilaka AAL, Panabokke RG. 1981. Studies on medicinal plants of Sri Lanka: occurence of pyrrolizidine alkaloids and hepatotoxic properties in some traditional medical herbs. Journal of Ethnopharmacology 4:159-177.
Banerjee TD, Ramos D, Monteiro A. 2020. Expression of Multiple engrailed Family Genes in Eyespots of Bicyclus anynana Butterflies Does Not Implicate the Duplication Events in the Evolution of This Morphological Novelty. Frontiers in Ecology and Evolution 8.
Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455-477.
Barny LA, Tasca JA, Sanchez HA, Smith CR, Koptur S, Livshultz T, Minbiole KPC. 2021. Chemotaxonomic investigation of Apocynaceae for retronecine-type pyrrolizidine alkaloids using HPLC-MS/MS. Phytochemistry 185:112662.
Barny LA, Tasca JA, Sanchez HA, Smith CR, Koptur S, Livshultz T, Minbiole KPC. 2022. Corrigendum to "Chemotaxonomic investigation of Apocynaceae for retronecine-type pyrrolizidine alkaloids using HPLC-MS/MS" [Phytochemistry 185 (2021) 112662]. Phytochemistry 202:113309.
Benderoth M, Textor S, Windsor AJ, Mitchell-Olds T, Gershenzon J, Kroymann J. 2006. Positive selection driving diversification in plant secondary metabolism. Proc Natl Acad Sci U S A 103:9118-9123.
Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456:53-59.
Bienz S, Detterbeck R, Ensch C, Guggisberg A, Häusermann U, Meisterhans C, Wendt B, Werner C, Hesse M. 2002. Putrescine, spermidine, spermine, and related polyamine alkaloids. Alkaloids Chem Biol 58:83-338.
Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120.

Bradford MM. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical Biochemistry 72:248-254.
Burzynski EA, Minbiole KPC, Livshultz T. 2015. New sources of lycopsamine-type pyrrolizidine alkaloids and their distribution in Apocynaceae. Biochemical Systematics and Ecology 59:331-339.
Cacho NI, Kliebenstein DJ, Strauss SY. 2015. Macroevolutionary patterns of glucosinolate defense and tests of defense-escalation and resource availability hypotheses. New Phytol 208:915-927.
Cameron M, Williams HE. 2007. Comparing compressed sequences for faster nucleotide BLAST searches. IEEE/ACM transactions on computational biology and bioinformatics.
Chen F, Ro DK, Petri J, Gershenzon J, Bohlmann J, Pichersky E, Tholl D. 2004. Characterization of a root-specific Arabidopsis terpene synthase responsible for the formation of the volatile monoterpene 1,8-cineole. Plant Physiol 135:1956-1966.
Chen KY, Liu AY-C. 1997. Biochemistry and function of hypusine formation on eukaryotic intitiation factor 5A. Biological Signals 6:105-109.
Coddington JA. 1988. Cladistic tests of adaptational hypotheses. Cladistics 4:3-22.
Colegate SM, Gardner DR, Betz JM, Fischer OW, Liede-Schumann S, Boppré M. 2016. Pro-toxic 1,2-dehydropyrrolizidine alkaloid esters, including unprecedented 10-membered macrocyclic diesters, in the medicinally-used *Alafia* cf. *caudata* and *Amphineurion marginatum* (Apocynaceae: Apocynoideae: Nerieae and Apocyneae). Phytochem Anal 27:257-276.
Cutter AD, Jovelin R. 2015. When natural selection gives gene function the cold shoulder. Bioessays 37:1169-1173.
Darwin C. 1859. The Origin of Species; And, the Descent of Man: Modern library.
del Castillo JB, de Aguirre AGE, Bretón JL, González AG, Trujillo J. 1970. Loroquin, a new necine isolated from *Urechites karwinsky* Mueller. Tetrahedron Letters 15:1219-1220.
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.
Dohm JC, Lottaz C, Borodina T, Himmelbauer H. 2008. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Res 36:e105.
Donoso A, Rivas C, Zamorano A, Pena A, Handford M, Aros D. 2021. Understanding Alstroemeria pallida Flower Colour: Links between Phenotype, Anthocyanins and Gene Expression. Plants (Basel) 10.
Edgar JA. 1984. Parsonsieae: ancestral larval foodplants of the Danainae and Ithomiinae. In: Vane-Wright RI, Ackery PR, editors. The biology of butterflies: Symposium of the Royal Entomological Society of London. p. 91-93.
Ehrlich PR, Raven PH. 1964. Butterflies and plants: a study in coevolution. Evolution 18:586-608.
Endress ME, Meve U, Middleton DJ, Liede-Schumann S. 2018. Apocynaceae. In: Kadereit JW, Bittrich V, editors. Flowering Plants. Eudicots: Apiales, Gentianales (except Rubiaceae). Berlin: Springer International Publishing.

Erb M, Reymond P. 2019. Molecular interactions between plants and insect herbivores. Annu Rev Plant Biol 70:1-31.
Erb M, Robert CA. 2016. Sequestration of plant secondary metabolites by insect herbivores: molecular mechanisms and ecological consequences. Curr Opin Insect Sci 14:8-11.
Espírito Santo FdSd, Rapini A, Ribeiro PL, Liede-Schumann S, Goyder DJ, Fontella-Pereira J. 2019. Phylogeny of the tribe Marsdenieae (Apocynaceae), reinstatement of *RueHSSia* and the taxonomic treatment of the genus in Brazil. Kew Bulletin 74.
Feng T, Yang Y, Busta L, Cahoon EB, Wang H, Lu S. 2019. FAD2 Gene Radiation and Positive Selection Contributed to Polyacetylene Metabolism Evolution in Campanulids. Plant Physiol 181:714-728.
Fishbein M, Livshultz T, Straub SCK, Simões AO, Boutte J, McDonnell A, Foote A. 2018. Evolution on the backbone: Apocynaceae phylogenomics and new perspectives on growth forms, flowers, and fruits. Am J Bot 105:495-513.
Force A, Lynch M, Pickett FB, Amores A, Yan Y-l, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531-1545.
Franke J, Kim J, Hamilton JP, Zhao D, Pham GM, Wiegert-Rininger K, Crisovan E, Newton L, Vaillancourt B, Tatsis E, et al. 2019. Gene Discovery in Gelsemium Highlights Conserved Gene Clusters in Monoterpene Indole Alkaloid Biosynthesis. Chembiochem 20:83-87.
Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. 2005. Protein Identification and Analysis Tools on the ExPASy Server. In. The Proteomics Protocols Handbook: Humana Press. p. 571–607.
Gill GP, Bryant CJ, Fokin M, Huege J, Fraser K, Jones C, Cao M, Faville MJ. 2018. Low pyrrolizidine alkaloid levels in perennial ryegrass is associated with the absence of a homospermidine synthase gene. BMC Plant Biol 18:56.
Gumulya Y, Gillam EM. 2017. Exploring the past and the future of protein evolution with ancestral sequence reconstruction: the 'retro' approach to protein engineering. Biochem J 474:1-19.
Halabi K, Karin EL, Guéguen L, Mayrose I. 2020. A codon model for associating phenotypic traits with altered selective patterns of sequence evolution. bioRxiv.
Harismendy O, Ng PC, Strausberg RL, Wang X, Stockwell TB, Beeson KY, Schork NJ, Murray SS, Topol EJ, Levy S, et al. 2009. Evaluation of next generation sequencing platforms for population targeted sequencing studies. Genome Biol 10:R32.
Hartmann T. 1999. Chemical ecology of pyrrolizidine alkaloids. Planta 207:483-495.
Hartmann T, Theuring C, Witte L, Pasteels JM. 2001. Sequestration, metabolism and partial synthesis of tertiary pyrrolizidine alkaloids by the neotropical leaf-beetle *Platyphora boucardi*. Insect Biochemistry and Molecular Biology 31:1041-1056.
Hartmann T, Witte L. 1995. Chemistry, biology and chemoecology of the pyrrolizidine alkaloids. In. Alkaloids: chemical and biological perspectives: Elsevier. p. 155-233.
Hoopes GM, Hamilton JP, Kim J, Zhao D, Wiegert-Rininger K, Crisovan E, Buell CR. 2018. Genome Assembly and Annotation of the Medicinal Plant Calotropis gigantea, a Producer of Anticancer and Antimalarial Cardenolides. G3 (Bethesda) 8:385-391.
Joly-Lopez Z, Flowers JM, Purugganan MD. 2016. Developing maps of fitness consequences for plant genomes. Curr Opin Plant Biol 30:101-107.

Kaltenegger E, Eich E, Ober D. 2013. Evolution of homospermidine synthase in the Convolvulaceae: a story of gene duplication, gene loss, and periods of various selection pressures. Plant Cell 25:1213-1227.
Kaltenegger E, Ober D. 2015. Paralogue interference affects the dynamics after gene duplication. Trends Plant Sci 20:814-821.
Kaltenegger E, Prakashrao AS, Cicek SS, Ober D. 2021. Development of an activity assay for characterizing deoxyhypusine synthase and its diverse reaction products. FEBS Open Bio.
Kapralov MV, Smith JA, Filatov DA. 2012. Rubisco evolution in C(4) eudicots: an analysis of Amaranthaceae sensu lato. PLoS One 7:e52974.
Katoh K, Misawa K, Kuma K-i, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059-3066.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772-780.
Kim KH, Lee KH, Choi SU, Kim KR, Lee KR. 2010. ChemInform Abstract: Pyrrolizidine Alkaloids from the Roots of Paris verticillata. ChemInform 41:no-no.
Kimura M. 1968a. Evolutionary rate at the molecular level. Nature 217:624-626.
Kimura M. 1968b. Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. Genet Res 11:247-269.
Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. 2006. Automated phylogenetic detection of recombination using a genetic algorithm. Mol Biol Evol 23:1891-1901.
Kreis W, Munkert J. 2019. Exploiting enzyme promiscuity to shape plant specialized metabolism. J Exp Bot 70:1435-1445.
Kroymann J. 2011. Natural diversity and adaptation in plant secondary metabolism. Curr Opin Plant Biol 14:246-251.
Liao D-I, Wolff EC, Park MH, Davies DR. 1998. Crystal structure of NAD complex of human deoxyhypysine synthase: an enzyme with a ball-and-chain mechanism for blocking the active site. Structure 6:23-32.
Liede-Schumann S, Reuss SJ, Meve U, Gâteblé G, Livshultz T, Forster PI, Wanntorp L, Rodda M. 2022. Phylogeny of Marsdenieae (Apocynaceae, Asclepiadoideae) based on chloroplast and nuclear loci, with a conspectus of the genera. Taxon.
Liu H, Naismith JH. 2008. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. BMC Biotechnol 8:91.
Livshultz T. 2010. The phylogenetic position of milkweeds (Apocynaceae subfamilies Secamonoideae and Asclepiadoideae): Evidence from the nucleus and chloroplast. Taxon 59:1016-1030.
Livshultz T, Kaltenegger E, Straub SCK, Weitemier K, Hirsch E, Koval K, Mema L, Liston A. 2018. Evolution of pyrrolizidine alkaloid biosynthesis in Apocynaceae: revisiting the defence de-escalation hypothesis. New Phytol 218:762-773.
Livshultz T, Middleton DJ, Endress ME, Williams JK. 2007. Phylogeny of Apocynoideae and the APSA Clade (Apocynaceae s.l.). Annals of the Missouri Botanical Garden 94:324-359.

McKain M, Wilson M. 2017. Fast-Plast: Rapid de novo assembly and finishing for whole chloroplast genomes. 2017. Zenodo. mrmckain/Fast-Plast: Fast-Plast v.1.2.6 (Version v.1.2.6).
Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees. Proceedings of the Gateway Computing Environments Workshop (GCE); 14 Nov. 2010; New Orleans. p. 1-8.
Moghe GD, Last RL. 2015. Something Old, Something New: Conserved Enzymes and the Evolution of Novelty in Plant Specialized Metabolism. Plant Physiol 169:1512-1523.
Moore AJ, Vos JM, Hancock LP, Goolsby E, Edwards EJ. 2018. Targeted Enrichment of Large Gene Families for Phylogenetic Inference: Phylogeny and Molecular Evolution of Photosynthesis Genes in the Portullugo Clade (Caryophyllales). Syst Biol 67:367-383.
Moore BD, Andrew RL, Kulheim C, Foley WJ. 2014. Explaining intraspecific diversity in plant secondary metabolites in an ecological context. New Phytol 201:733-750.
Morales JF. 2017. Systematics of the tribe Echiteae and the genus *Prestonia* (Apocynaceae, Apocynoideae). [Universität Bayreuth.
Morales JF, Endress ME, Liede-Schumann S. 2017. Sex, drugs and pupusas: Disentangling relationships in Echiteae (Apocynaceae). Taxon 66:623-644.
Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. 2012. Detecting individual sites subject to episodic diversifying selection. PLoS Genet 8:e1002764.
Nawaz HR, Malik A, Muhammad P, Ahmed S, Riaz M. 2000. Chemical constituents of *Ajuga parviflora*. Zeitschrift für Naturforschung B 55:100-103.
Noda-Garcia L, Tawfik DS. 2020. Enzyme evolution in natural products biosynthesis: target- or diversity-oriented? Curr Opin Chem Biol 59:147-154.
Ober D. 2010. Gene duplications and the time thereafter - examples from plant secondary metabolism. Plant Biol (Stuttg) 12:570-577.
Ober D, Gibas L, Witte L, Hartmann T. 2003. Evidence for general occurence of homosperminidine in plants and its supposed origin as by-product of deoxyhypusine. Phytochemistry 62:339-344.
Ober D, Harms R, Witte L, Hartmann T. 2003. Molecular evolution by change of function. Alkaloid-specific homospermidine synthase retained all properties of deoxyhypusine synthase except binding the eIF5A precursor protein. J Biol Chem 278:12805-12812.
Ober D, Hartmann T. 1999a. Deoxyhypusine synthase from tobacco: cDNA isolation, characterization, and bacterial expression of an enzyme with extended substrate specificity. Journal of Biological Chemistry 274:32040-32047.
Ober D, Hartmann T. 1999b. Homospermidine synthase, the first pathway-specific enzyme of pyrrolizidine alkaloid biosynthesis, evolved from deoxyhypusine synthase. Proceedings of the National Academy of Sciences 96:14777-14782.
Ober D, Hartmann T. 2000. Phylogenetic origin of a secondary pathway: the case of pyrrolizidine alkaloids. Plant Molecular Biology 44:445-450.
Ober D, Kaltenegger E. 2009. Pyrrolizidine alkaloid biosynthesis, evolution of a pathway in plant secondary metabolism. Phytochemistry 70:1687-1695.
Olson-Manning CF, Wagner MR, Mitchell-Olds T. 2012. Adaptive evolution: evaluating empirical support for theoretical predictions. Nat Rev Genet 13:867-877.

Panchy N, Lehti-Shiu M, Shiu S-H. 2016. Evolution of Gene Duplication in Plants. Plant Physiol 171:2294-2316.
Park MH, Nishimura K, Zanelli CF, Valentini SR. 2010. Functional significance of eIF5A and its hypusine modification in eukaryotes. Amino Acids 38:491-500.
Parto S, Lartillot N. 2018. Molecular adaptation in Rubisco: Discriminating between convergent evolution and positive selection using mechanistic and classical codon models. PLoS One 13:e0192697.
Pasteels JM, Termonia A, Windsor DM, Witte L, Theuring C, Hartmann T. (Pasteels2001 co-authors). 2001. Pyrrolizidine alkaloids and pentacyclic triterpene saponins in the defensive secretions of *Platyphora* leaf beetles. Chemoecology 11:113-120.
Piatkowski BT, Imwattana K, Tripp EA, Weston DJ, Healey A, Schmutz J, Shaw AJ. 2020. Phylogenomics reveals convergent evolution of red-violet coloration in land plants and the origins of the anthocyanin biosynthetic pathway. Mol Phylogenet Evol 151:106904.
Pichersky E, Gang DR. 2000. Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. Trends Plant Sci 5:439-445.
Pichersky E, Lewinsohn E. 2011. Convergent evolution in plant specialized metabolism. Annu Rev Plant Biol 62:549-566.
Prakashrao AS, Beuerle T, Simões ARG, Hopf C, Çiçek SS, Stegemann T, Ober D, Kaltenegger E. 2022. The long road of functional recruitment—The evolution of a gene duplicate to pyrrolizidine alkaloid biosynthesis in the morning glories (Convolvulaceae). Plant Direct 6:e420.
Rausher MD. 2013. The evolution of genes in branched metabolic pathways. Evolution 67:34-48.
Rausher MD, Delph LF. 2015. Commentary: When does understanding phenotypic evolution require identification of the underlying genes? Evolution 69:1655-1664.
Rausher MD, Huang J. 2016. Prolonged Adaptive Evolution of a Defensive Gene in the Solanaceae. Mol Biol Evol 33:143-151.
Reeve HK, Sherman PW. 1993. Adaptation and the goals of evolutionary research. The Quarterly Review of Biology 68:1-32.
Reimann A, Nurhayati N, Backenköhler A, Ober D. 2004. Repeated evolution of the pyrrolizidine alkaloid-mediated defense system in separate angiosperm lineages. Plant Cell 16:2772-2784.
Rey C, Lanore V, Veber P, Gueguen L, Lartillot N, Semon M, Boussau B. 2019. Detecting adaptive convergent amino acid evolution. Philos Trans R Soc Lond B Biol Sci 374:20180234.
Rinker DC, Specian NK, Zhao S, Gibbons JG. 2019. Polar bear evolution is marked by rapid changes in gene copy number in response to dietary shift. Proc Natl Acad Sci U S A 116:13446-13451.
Sadier A, Davies KTJ, Yohe LR, Yun K, Donat P, Hedrick BP, Dumont ER, Davalos LM, Rossiter SJ, Sears KE. 2018. Multifactorial processes underlie parallel opsin loss in neotropical bats. Elife 7.
Sasaki K, Hirata Y. 1970. The alkaloids of Anodendron affine Druce. Tetrahedron Letters 26:2119-2126.

Scossa F, Fernie AR. 2020. The evolution of metabolism: How to test evolutionary hypotheses at the genomic level. Comput Struct Biotechnol J 18:482-500.
Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL. 2015. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Mol Biol Evol 32:1342-1353.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312-1313.
Straub SC, Moore MJ, Soltis PS, Soltis DE, Liston A, Livshultz T. 2014. Phylogenetic signal detection from an ancient rapid radiation: Effects of noise reduction, long-branch attraction, and model selection in crown clade Apocynaceae. Mol Phylogenet Evol 80:169-185.
Straub SC, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A. 2012. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. Am J Bot 99:349-364.
Straub SCK, Boutte J, Fishbein M, Livshultz T. 2020. Enabling evolutionary studies at multiple scales in Apocynaceae through Hyb-Seq. Appl Plant Sci 8:e11400.
Tamariz J, Burgueno-Tapia E, Vazquez MA, Delgado F. 2018. Pyrrolizidine Alkaloids. Alkaloids Chem Biol 80:1-314.
Tasca JA, Smith CR, Burzynski EA, Sundberg BN, Lagalante AF, Livshultz T, Minbiole KPC. 2018. HPLC-MS detection of pyrrolizidine alkaloids and their *N*-oxides in herbarium specimens dating back to the 1850s. Applications in Plant Sciences 6:e1143.
Umland TC, Wolff EC, Park MH, Davies DR. 2004. A new crystal structure of deoxyhypusine synthase reveals the configuration of the active enzyme and of an enzyme.NAD.inhibitor ternary complex. J Biol Chem 279:28697-28705.
Wator E, Wilk P, Grudnik P. 2020. Half Way to Hypusine-Structural Basis for Substrate Recognition by Human Deoxyhypusine Synthase. Biomolecules 10.
Weiss KM. 2010. Seeing the forest through the gene-trees. Evolutionary Anthropology: Issues, News, and Reviews 19:210-221.
Weiss KM, Fullerton SM. 2000. Phenogenetic drift and the evolution of genotype-phenotype relationships. Theor Popul Biol 57:187-195.
Weng JK, Philippe RN, Noel JP. 2012. The rise of chemodiversity in plants. Science 336:1667-1670.
Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. 2015. RELAX: detecting relaxed selection in a phylogenetic framework. Mol Biol Evol 32:820-832.
Whitfield JB, Lockhart PJ. 2007. Deciphering ancient rapid radiations. Trends Ecol Evol 22:258-265.
Whittall JB, Syring J, Parks M, Buenrostro J, Dick C, Liston A, Cronn R. 2010. Finding a (pine) needle in a haystack: chloroplast genome sequence divergence in rare and widespread pines. Mol Ecol 19 Suppl 1:100-114.
Wicke S, Muller KF, dePamphilis CW, Quandt D, Bellot S, Schneeweiss GM. 2016. Mechanistic model of evolutionary rate variation en route to a nonphotosynthetic lifestyle in plants. Proc Natl Acad Sci U S A 113:9045-9050.
Wink M. 2011. Biochemistry of plant secondary metabolism. London: John Wiley & Sons.

Wisecaver JH, Borowsky AT, Tzin V, Jander G, Kliebenstein DJ, Rokas A. 2017. A Global Coexpression Network Approach for Connecting Genes to Specialized Metabolic Pathways in Plants. Plant Cell 29:944-959.
Wright KM, Rausher MD. 2010. The evolution of control and distribution of adaptive mutations in a metabolic pathway. Genetics 184:483-502.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586-1591.
Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Cabios Applications 13:555-556.
Yang Z, **Wang T**. 1995. Mixed model analysis of DNA sequence evolution. Biometrics 51:552-561.
Zakaria MM, Stegemann T, Sievert C, Kruse LH, Kaltenegger E, Girreser U, Cicek SS, Nimtz M, Ober D. 2022. Insights into polyamine metabolism: homospermidine is double-oxidized in two discrete steps by a single copper-containing amine oxidase in pyrrolizidine alkaloid biosynthesis. Plant Cell.

Figure Legends

Fig. 1: Function of deoxyhypusine synthase (DHS, a) and homospermidine synthase (HSS, b). **a.** DHS activates the eIF5A translation elongation factor by transferring the amino-butyl moiety of spermidine to a lysine residue on the eIF5A precursor protein to form deoxyhypusine (Chen and Liu 1997). DHS can also transfer the amino-butyl moiety of spermidine to putrescine to form homospermidine (Ober and Hartmann 1999a). **b.** HSS (homospermidine synthase) has lost activity with eIF5A, but has maintained and optimized activity with putrescine, producing an order of magnitude more homospermidine than DHS (Ober, Harms, et al. 2003; Reimann, et al. 2004; Kaltenegger, et al. 2013). Homospermidine is precursor to the necine base of a pyrrolizidine alkaloid.

Fig. 2: Phylogenetic relationships among Apocynaceae subfamilies and tribes, with pyrrolizidine alkaloid (PA) distribution (orange circles), duplications that gave rise to *HSS*-like genes (stars), and distribution of motifs and pseudogenes (Ψ) among *HSS*-like genes mapped. Topology reflects statistically significant relationships supported by the polytomy test in the nuclear ASTRAL analysis of Antonelli et al. (2021), with addition of Odontadenieae,

placed here in a polytomy with Mesechiteae and Echiteae (Livshultz, et al. 2007; Livshultz 2010; Morales, et al. 2017; Fishbein, et al. 2018). Stars at nodes represent reconstructions of the *DHS/HSS*-like gene duplication. Livshultz, et al. (2018) reconstructed this duplication after the divergence of Wrightieae (white star).This study places it in the MRCA of the APSA clade (black star). Orange circles at tips represent the presence of PAs with high confidence as defined in Barny, et al. (2021) in at least one species of the group. The amino acid motif (positions 269 and 273) of *HSS*-like genes are indicated at tips (VXXXD in red, IXXXD in blue, and IXXXN in green). Nerieae has multiple motifs (VXXXD in *Strophanthus* and *Isonema*, IXXXD in *Alafia*; Marsdenieae has two *HSS*-like paralogs with IXXXD and IXXXN motifs (see Fig. 3). Ψ indicates that at least one species in the lineage has only a pseudogenized copy of *HSS*.

Fig. 3: Summary of *DHS/HSS* gene trees based on the reduced dataset topology (subpanels **a, b**, see Fig. 4 for entire tree) and the full dataset topology (subpanels **c, d**, see Fig. S3 for entire tree) with branches colored according to the reconstructed I269VXXXN273D motif (green: IXXXN, blue: IXXXD, red: VXXXD). Distribution of pyrrolizidine alkaloids indicated with orange circles next to the *HSS* gene lineage of the corresponding taxon. The two alternative evolutionary hypotheses of PA biosynthesis and HSS optimization: single origin of PAs followed by multiple independent losses (defense de-escalation, subpanels **a, c**) and multiple parallel origins (defense escalation, subpanels **b, d**) are illustrated on each gene tree topology via the patterns of selection [nonsynonymous/synonymous substitution ratios (ω)] on *HSS* predicted under each scenario. Branches and clades marked by letters (A-L) are used in tests for selection (see Tables 1 and 2). Under the single origin/defense de-escalation model (subpanels a, c),

PA biosynthesis evolved in the MRCA of all PA-producers, and the ancestral HSS enzyme was optimized under positive selection (Branch A, ω>1). Subsequent parallel losses of PA biosynthesis, under negative or neutral selection, result in neutral evolution of the *HSS* locus (ω=1) and evolution of a more *DHS*-like from a more *HSS*-like motif, i.e. VXXXD to IXXXD in the ancestral *HSS*-like gene of *Alafia* (Branch K, subpanel a) and IXXXD to IXXXN in the ancestral *HSS*-like gene of Asclepiadeae (Branch I, subpanels a and c) and in the IXXXN *HSS*-like paralog of Marsdenieae (Branch J, subpanels a and c). Relaxation of selection in groups marked by K and I and J is tested relative to reference groups marked by G and L respectively (see *Relax: tests for shifts in selection intensity* in Results and Table 2). The parallel origins/defense escalation model (subpanels b, d) predicts that the ancestral *HSS*-like locus (Branch A) was under neutral or purifying selection (ω<=1) followed by multiple independent origins of PAs, and independent optimization of HSS function under positive selection (ω>1) in the ancestral *HSS* loci of PA-producing clades: Echiteae (Branch C) and within Apocyneae (Branches D and E, subpanels b and d), and/or in *HSS* loci in which the PA-associated VXXXD motif evolved. Evolution of the VXXXD motif is reconstructed either in the ancestral HSS locus of Nerieae and Malouetieae (Branch G, subpanel b) or independently in the ancestral HSS loci of *Isonema*, *Strophanthus*, and Malouetieae (Branch F) (subpanel d). VXXXD is also reconstructed as evolving in the ancestral HSS of Apocyneae and MOE (Branch B) in both trees (subpanels b and d). Asterisks at nodes represent bootstrap support > 90%.

Fig. 4a: *DHS/HSS* gene tree from reduced dataset (see Methods: *Tree Construction*, Fig. 2 for species tree, and Figs. 3a, b for summary gene tree). **a.** *DHS*-like sequences from

outgroup and rauvolfioid taxa. Branch lengths are proportional to probability of nucleotide substitution. Branch widths are proportional to the relative rate of nonsynonymous to synonymous amino acid changes (ω), divided into four categories: 1) $\omega<0.7$, purifying selection; 2) $0.7< \omega < 1.0$, neutral evolution; 3) $1< \omega < 2.5$, slightly positive selection; 4) $\omega >2.5$, strongly positive selection. Branch color depicts the selection intensity parameter, k, calculated under the RELAX descriptive model (see Methods: *Tests for selection*) , with darkest blue corresponding to k=0, i.e. $\omega_1, \omega_2, \omega_3= 1$, complete relaxation of selection, and darkest red to k>=2.5, i.e. $\omega_1<<1$, $\omega_2=1$, $\omega_3>>1$, intensification of selection. Substitutions at amino acid positions experiencing episodic diversifying selection (see Table 3) and in the I269VXXXN273D motif are mapped on the branches. Branches and lineages A-L, used in selection tests, (see Fig. 3, Tables 1-3) are labeled with letters and arrowheads. Presence of PAs is indicated next to all sequences from that species (i.e. both *DHS*- and *HSS*-like sequences): circles correspond to PA presence in the species with high confidence, as defined by Barny, et al. (2021); triangles correspond to PA presence in the genus with high confidence. Bootstrap support > 90 % is indicated with asterisks.

Fig. 4b: *DHS/HSS* gene tree from reduced dataset (see Methods: *Tree Construction*, Fig. 2 for species tree, Figs. 3a, b for summary gene tree, legend for Fig. 4a for detailed description). **b.** *DHS*-like sequences from APSA clade taxa, excluding Wrightieae *DHS2*.

Fig. 4c: *DHS/HSS* gene tree from reduced dataset (see Methods: *Tree Construction*, Fig. 2 for species tree, Figs. 3a, b for summary gene tree, legend for Fig. 4a for detailed description). **c**. Wrightieae *DHS2* and Nerieae and Malouetieae *HSS*-like sequences.

Fig. 4d: *DHS/HSS* gene tree from reduced dataset (see Methods: *Tree Construction*, Fig. 2 for species tree, Figs. 3a, b for summary gene tree, legend for Fig. 4a for detailed description). **d.** *HSS*-like sequences from Rhabdadenieae, Apocyneae, MOE, and Periplocoideae.

Fig. 4e: *DHS/HSS* gene tree from reduced dataset (see Methods: *Tree Construction*, Fig. 2 for species tree, Figs. 3a, b for summary gene tree, legend for Fig. 4a for detailed description). **e.** HSS-like sequences from Baisseeae, Secamonoideae, Fockeeae, Asclepiadeae, and Marsdenieae.

Fig. S1: Maximum likelihood tree from initial alignment (see Methods: *Gene assembly from contigs*, Table S1), including all assembled contigs and previously sequenced *DHS*-like and *HSS*-like genes. Arrows link contigs concatenated for downstream analyses. Contigs with non-terminal stop codons marked by "Ψ" in the sequence name. Bootstrap support values >50% mapped at nodes.

Fig. S2: Maximum likelihood tree from full dataset alignment (see Methods: *Gene assembly from contigs, Construction of matrices*, Table S1), including all fully assembled gene sequences, concatenated contigs, additional short contigs, potential pseudogene contigs, and previously sequenced *DHS*-like and *HSS*-like genes. Contigs with non-terminal stop codons marked by "Ψ" in the sequence name. Boot support values >50% mapped at nodes.

Fig. S3: Maximum likelihood tree, full dataset topology [see Methods: *Gene Tree Construction*: reduced dataset alignment, constrained to the tree topology from the full dataset alignment (Fig. S2)]. Branch lengths are proportional to probability of nucleotide

substitution. Branch widths are proportional to the relative rate of nonsynonymous to synonymous amino acid changes (ω), divided into four categories: 1) $\omega<0.7$, purifying selection; 2) $0.7< \omega < 1.0$, neutral evolution; 3) $1< \omega < 2.5$, slightly positive selection; 4) $\omega >2.5$, strongly positive selection. Branch color depicts the selection intensity parameter, k, calculated under the RELAX descriptive model (see Methods: *Tests for selection*) , with darkest blue corresponding to k=0, i.e. $\omega_1, \omega_2, \omega_3= 1$, complete relaxation of selection, and darkest red to k>=2.5, i.e. $\omega_1<<1$, $\omega_2=1$, $\omega_3>>1$, intensification of selection. Substitutions at amino acid positions experiencing episodic diversifying selection (see Table 3) and in the I269VXXXN273D motif are mapped on the branches. Branches and lineages A-L, used in selection tests, (see Fig. 3, Tables 1-3) are labeled with letters and arrowheads. Presence of PAs is indicated next to all sequences from that species (i.e. both *DHS*- and *HSS*-like sequences): circles correspond to PA presence in the species with high confidence, as defined by Barny, et al. (2021); triangles correspond to PA presence in the genus with high confidence. Bootstrap support > 90 % is indicated with asterisks.

Fig. S4: Identification of recombinants among *HSS*-like sequences from Marsdenieae. **a**. ML topology (See Methods: *Gene tree construction* and *Tests for recombination*) of the full Marsdenieae *HSS*-like sequences showing anomalous placement of the *Mardenia tinctoria* sequence (IXXXD motif) in the IXXXN clade and the *Gongronemopsis truncata* sequence (IXXXN motif) among IXXXD sequences. **b.** ML topology with the 5' and 3' ends of the *M. tinctoria* and *G. truncata* sequences included as distinct terminals, separated at a likely recombination breakpoint identified with GARD (See Methods: *Tests for recombination*). The *Gongronemopsis truncata* sequence is likely a

recombinant between a sequence in the IXXXN clade and the IXXXD grade, but the anomalous position of the 3' end with the IXXXN motif is not caused by recombination. the Bootstrap values, displayed at nodes and amino acid motif at positions 269 and 273 are noted next to sequence names.

Fig. S5: **a**. Confirmation via gel eletrophoresis of *Parsonsia alboflavescens* HSS enzyme (wild-type and mutants) expression and purification. **b.** Confirmation via HPLC (high performance liquid chromatography) that the enzymes form homotetramers (bottom).

Fig. S6: Kinetic curves of *Parsonsia alboflavescens* DHS and HSS wild-type activity with eIF5A.

Fig. S7: Full dataset topology tree with node numbers referenced in the PAML reconstruction of ancestral sequences and in Tables S6 and S9.

Fig. S8: Reduced dataset tree with node numbers referenced in the PAML reconstruction of ancestral sequences and in Tables S7 and S10.

Tables

Table 1:
Tests for positive selection on a subset of sites on selected branches in the full dataset and reduced dataset topologies with aBSREL. Branches lettered in Figs. 3 (both topologies), 4 (Reduced dataset topology) and S3 (Full dataset topology); "--" indicates that the branch does not occur in the topology. The likelihood ratio test (LRT) compares the likelihood of the null (ω <1) versus the alternative model (optimized ω). The optimized model fits up to two ω classes to these data, with >97% of sites evolving under ω1 and <3% of sites evolving under ω2. LRT=0 indicates no improvement in model fit and LRT >>0 indicates better model fit, significance indicated by p-values (raw and corrected for multiple testing). Significant p-values indicated with asterisk.

Full Dataset Toplogy Branch	**Branch**	**LRT**	**ω1**	**% sites**	**ω2**	**% sites**	**raw p-value**	**corrected p-value**
Ancestral *HSS*-like (excl. Wrightieae)	A	0	0.821	100	--	--	1	1
MRCA Apocyneae+MOE *HSS*-like	B	0	0.235	100	--	--	1	1
MRCA Echiteae *HSS*-like	C	5.8055	0.248	97.7	18.7	2.3	0.0197*	0.1367
MRCA *Anodendron HSS*-like	D	0	0.133	100	--	--	1	1
MRCA *Anodendron+Amphineurion HSS-like*	E	0	0	100	--	--	1	1
MRCA Malouetieae *HSS*-like	F	11.1102	0	98.6	158	1.4	0.0013*	0.0133*
MRCA Nerieae+Malouetieae *HSS*-like	--	--	--	--	--	--	--	--
Amphineurion marginatum HSS-like		0	0.0515	100	--	--	1	1
Isonema smeathmannii HSS-like		8.0944	0.0958	98.5	197	1.5	0.0061*	0.0552
Marsdenia tinctoria HSS-like		0	0.353	100	--	--	1	1
Strophanthus boivinii HSS-like		7.8189	0.0843	97.5	16.5	2.5	0.007*	0.0564
Reduced Dataset Topology Branch	**Node**	**LRT**	**ω1**	**% sites**	**ω2**	**% sites**	**raw p-value**	**corrected p-value**
Ancestral *HSS*-like (excl. Wrightieae)	A	0	0.826	100	--	--	1	1
MRCA Apocyneae+MOE *HSS*-like	B	0	0.168	100	--	--	1	1
MRCA Echiteae *HSS*-like	C	5.5263	0.263	97.8	19.7	2.2	0.0227*	0.1814
MRCA *Anodendron HSS*-like	D	0	0.154	100	--	--	1	1
MRCA *Anodendron+Amphineurion HSS-like*	E	0	0	100	--	--	1	1
MRCA Malouetieae *HSS*-like	F	--	0	98.8	98	1.2	--	--
MRCA Nerieae+Malouetieae *HSS*-like	G	0	0.581	100	--	--	1	1

Amphineurion marginatum HSS-like	0	0.0515	100	--	--	1	1
Isonema smeathmannii HSS-like	--	0.0867	98.5	205	1.5	--	--
Marsdenia tinctoria HSS-like	0	0.353	100	--	--	1	1
Strophanthus boivinii HSS-like	--	0.075	97.3	15.6	2.7	--	--

Table 2:
Tests for significant shifts in selection strength on designated **branches with RELAX. Branches lettered** as in **Figs. 3 (both topologies), 4 (reduced dataset topology) and S3 (full dataset topology). The test branches (i.e. clades subtended by and including the lettered branches) are compared against the reference branches. RELAX estimates k as a selection intensity parameter. The null model fits three ω rate categories to the test and reference branches, and the alternative model raises the rates on the test branches to the power of k, ω^k. The LRT compares the null and alternative models.** A significant k>1 indicates that selection has intensified in test branches relative to reference branches, and a significant k<1indicates relaxation. Significant p-values (<0.05) indicated by an asterisk. Replicate runs of the D273N test [D273N (1) and D273N (2)] produced inconsistent results.

Comparison	**Test branches**	**Reference branches**	**Full dataset (Fig. S3)**			**Reduced dataset (Fig. 4)**		
			k	***LRT***	***p-value***	***k***	***LRT***	***p-value***
HSS vs. DHS	Clade A	Clade H	1.31	4.19	0.041*	1.31	2.81	0.094
IXXXN vs. IXXXD and VXXXD	Clade I, Clade J	remainder of Clade A	0.79	1.72	0.19	1.5	5.32	0.021*
V269I	Clade K	remainder of Clade G	--	--	--	1.17	0.14	0.711
D273N (1)	Clade I, Clade J	remainder of Clade L	0.87	2.85	0.091	0.86	3.65	0.056
D273N (2)	Clade I, Clade J	remainder of Clade L	1.2	10.4	0.0013*	0.86	3.93	0.047*

Table 3.
Tests for positive selection on individual amino acid positions on designated branches with MEME. Lettered branches A-G, as well as the *hss*-like *Amphineurion marginatum*, *Isonema smeathmannii*, *Marsdenia tinctoria*, and *Strophanthus boivinii* branches (Fig. 3) are tested in the reduced dataset topology (Fig. 4) and the full dataset topology (Fig. S3). The LRT compares the null model, where nonsynonymous to synonymous mutation ratios are restricted to $\omega < 1$, to the alternative model, where ω is unrestricted. Only amino acid (AA) positions that experienced significant positive selection in at least one of the tree topologies are reported; substitutions at these sites are mapped on the tested tree topologies (Figs. 4, S3). Significant p-values indicated with an asterisk. Possible function is inferred from homologous human DHS amino acid positions.

AA Position	Full dataset tree topology		Reduced dataset tree		Note	Possible function
	LRT	p-value	LRT	p-value		
19	4.7	0.04*	4.29	0.05		Ball and chain motif
41	5.04	0.04*	0	0.67		
51	6.27	0.02*	0	0.67		
75	7.22	0.01*	0.63	0.4	Synapomorphy for all *HSS*-like orthologs (excluding Wrightieae)	
200	4.22	0.06	0	0.67		
246	5.24	0.03*	1.55	0.23	100% conserved in *DHS*-like orthologs, including both Wrightieae paralogs	active site tunnel entrance
269	0	0.67	5.36	0.03*	part of I269VXXXN273D amino acid motif	active site tunnel entrance, interaction between monomers
275	9.77	0*	0	0.67		active site tunnel entrance

Table 4. Mean specific activities of wild-type and mutant *Parsonsia alboflavescens* HSS. For the HSS function assay (substrate=putrescine), the specific activity of homospermidine (hspd) production, as well as 1,3-diaminopropane (dap) was measured. Both DHS function assays (substrate=eIF5A and substrate=spermidine [spd]) measured production of dap and canavalmine (Can), Values in pkat mg^{-1} and the relative standard deviation of three replicates are reported in parentheses.

	HSS assay Aminobutyl acc. Putrescine		**DHS assay** Aminobutyl acc. eIF5A		**DHS assay -eIF5A** Aminobutyl acc. spd	
PaHSS enzyme	dap	hspd	dap	Can	dap	Can
Wildtype (VXXXD)	1307 (22%)	1317 (16%)	34 (3%)	13 (27%)	43 (14%)	12 (8%)
IXXXD	1122 (5%)	1145 (3 %)	95 (5%)	7 (2%)	82 (12%)	7 (4%)
VXXXN	429 (5%)	471 (1 %)	41 (7%) *	20 (8%) *	50 (5%)	30 (11%)
IXXXN	532 (5%)	526 (7%)	27 (3%) *	18 (4%) *	17 (19%)	11 (7%)

* deoxyhypusine was detected in the *in vitro* assay, after hydrolysis.

Figure 1

a
eIF5A
precursor
protein
spermidine +
DHS
putrescine
+ spermidine
eIF5A*
homospermidine
b
eIF5A
precursor
protein
spermidine +
HSS
putrescine
+ spermidine
eIF5A*
homospermidine

Figure 2

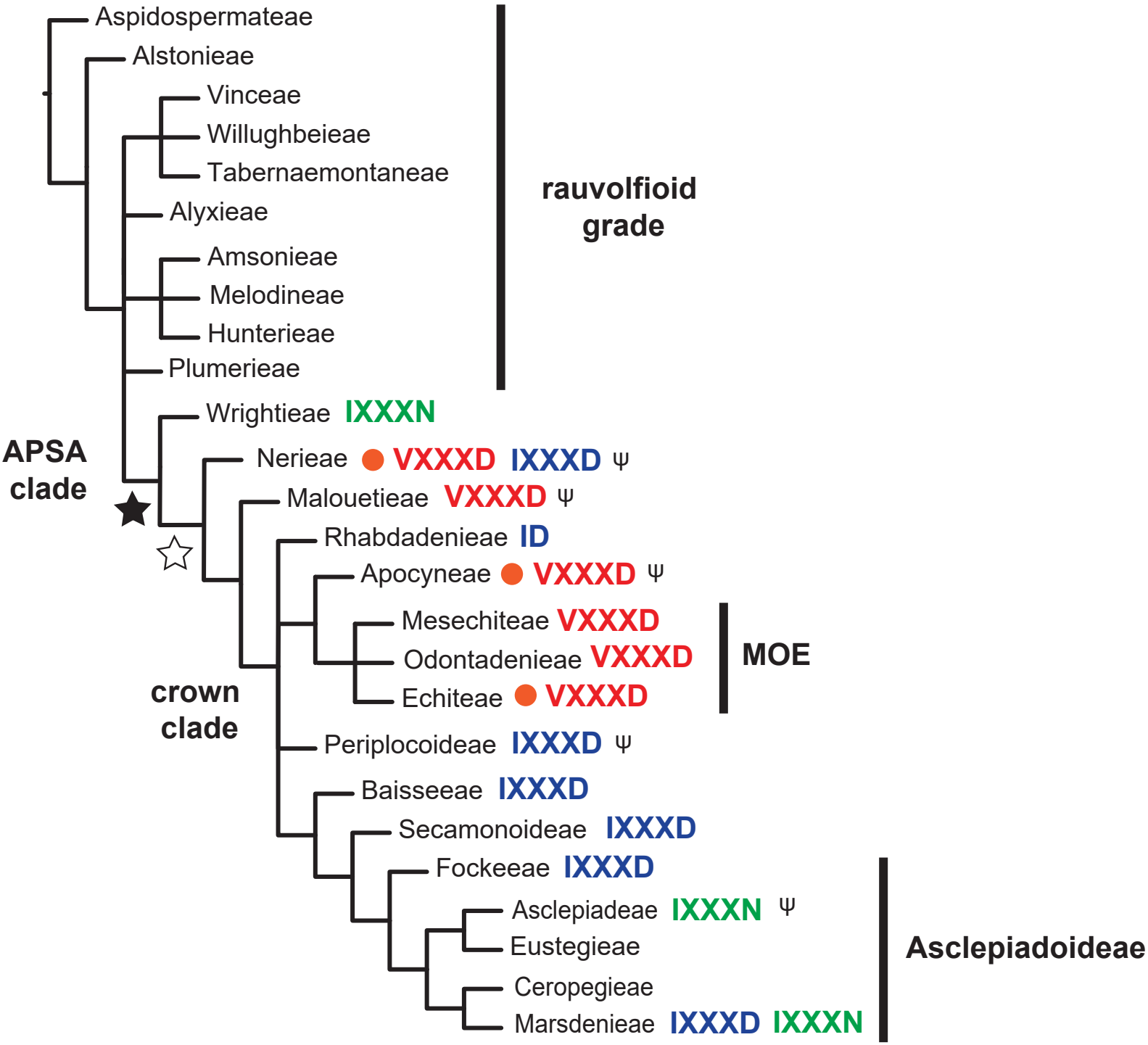
Aspidospermateae
Alstonieae
Vinceae
Willughbeieae
Tabernaemontaneae
Alyxieae
Amsonieae
Melodineae
Hunterieae
Plumerieae
rauvolfioid grade
Wrightieae IXXXN
APSA clade
Nerieae VXXXD IXXXD ψ
Malouetieae VXXXD ψ
Rhabdadenieae ID
Apocyneae VXXXD ψ
Mesechiteae VXXXD
Odontadenieae VXXXD
Echiteae VXXXD
MOE
crown clade
Periplocoideae IXXXD ψ
Baisseeae IXXXD
Secamonoideae IXXXD
Fockeeae IXXXD
Asclepiadeae IXXXN ψ
Eustegieae
Ceropegieae
Marsdenieae IXXXD IXXXN
Asclepiadoideae

Figure 3

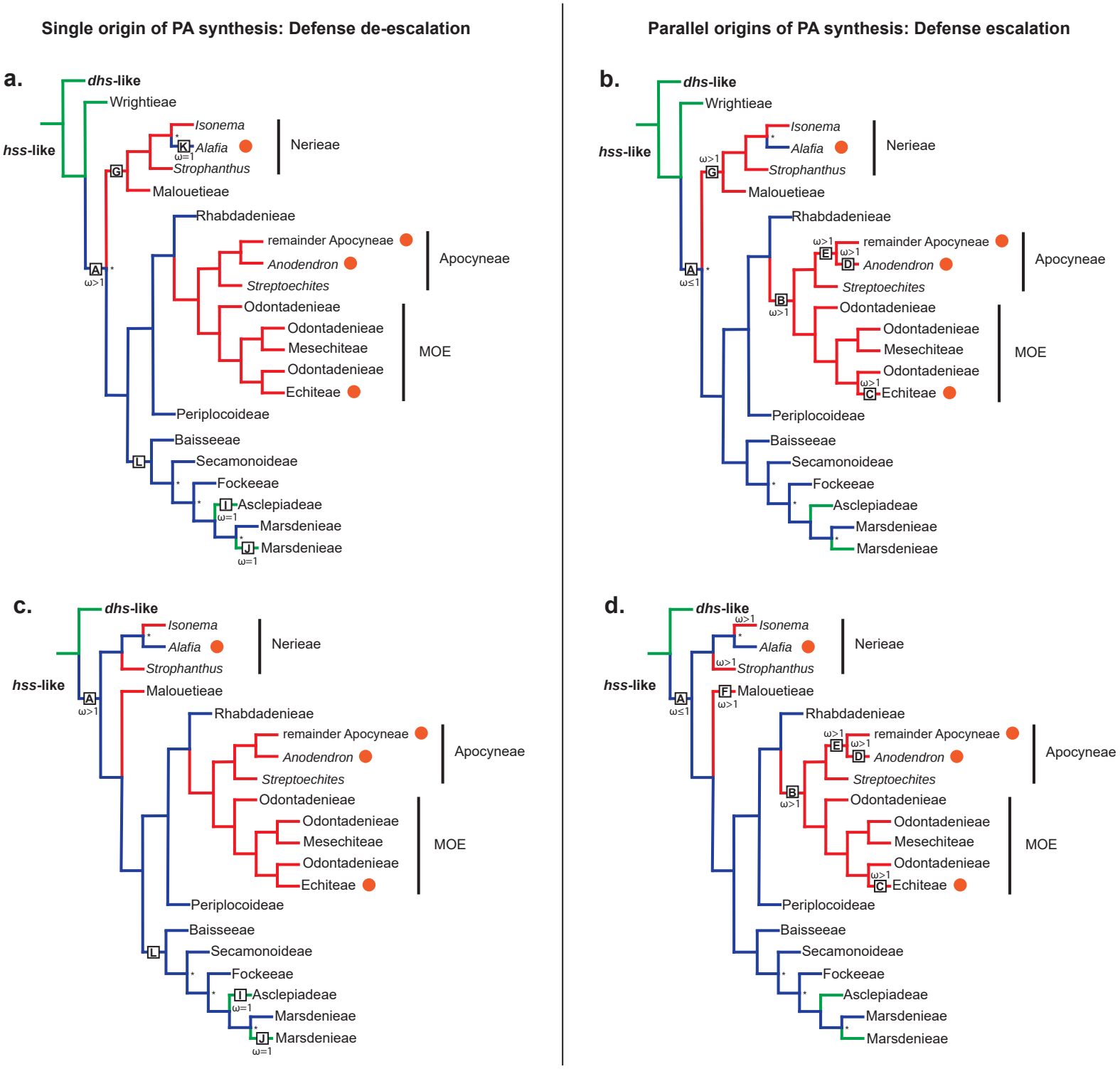
Single origin of PA synthesis: Defense de-escalation
Parallel origins of PA synthesis: Defense escalation
a.
b.
c.
d.
dhs-like
hss-like
Wrightieae
Isonema
Alafia
Strophanthus
Nerieae
Malouetieae
Rhabdadenieae
remainder Apocyneae
Anodendron
Streptoechites
Apocyneae
Odontadenieae
Mesechiteae
Echiteae
MOE
Periplocoideae
Baisseeae
Secamonoideae
Fockeeae
Asclepiadeae
Marsdenieae
A
B
C
D
E
F
G
I
J
K
L
ω>1
ω≤1
ω=1
*

Figure 4a

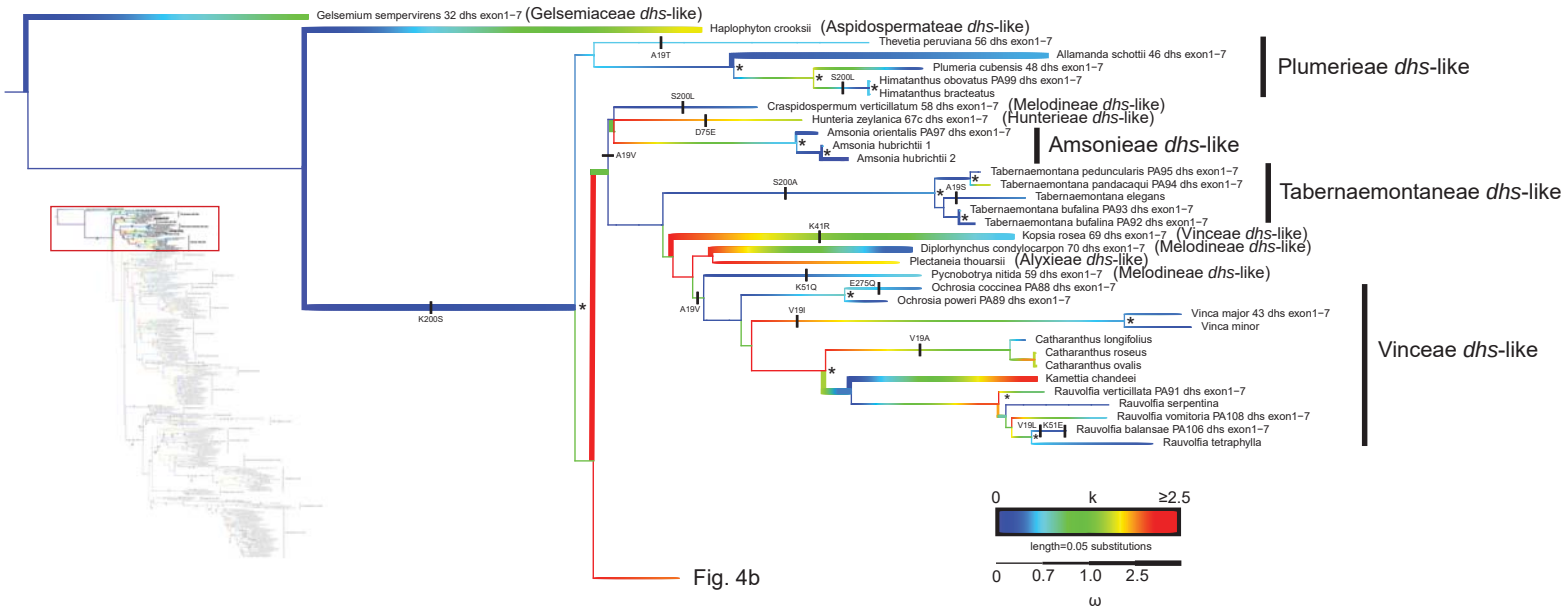

Fig. 4b

Figure 4b

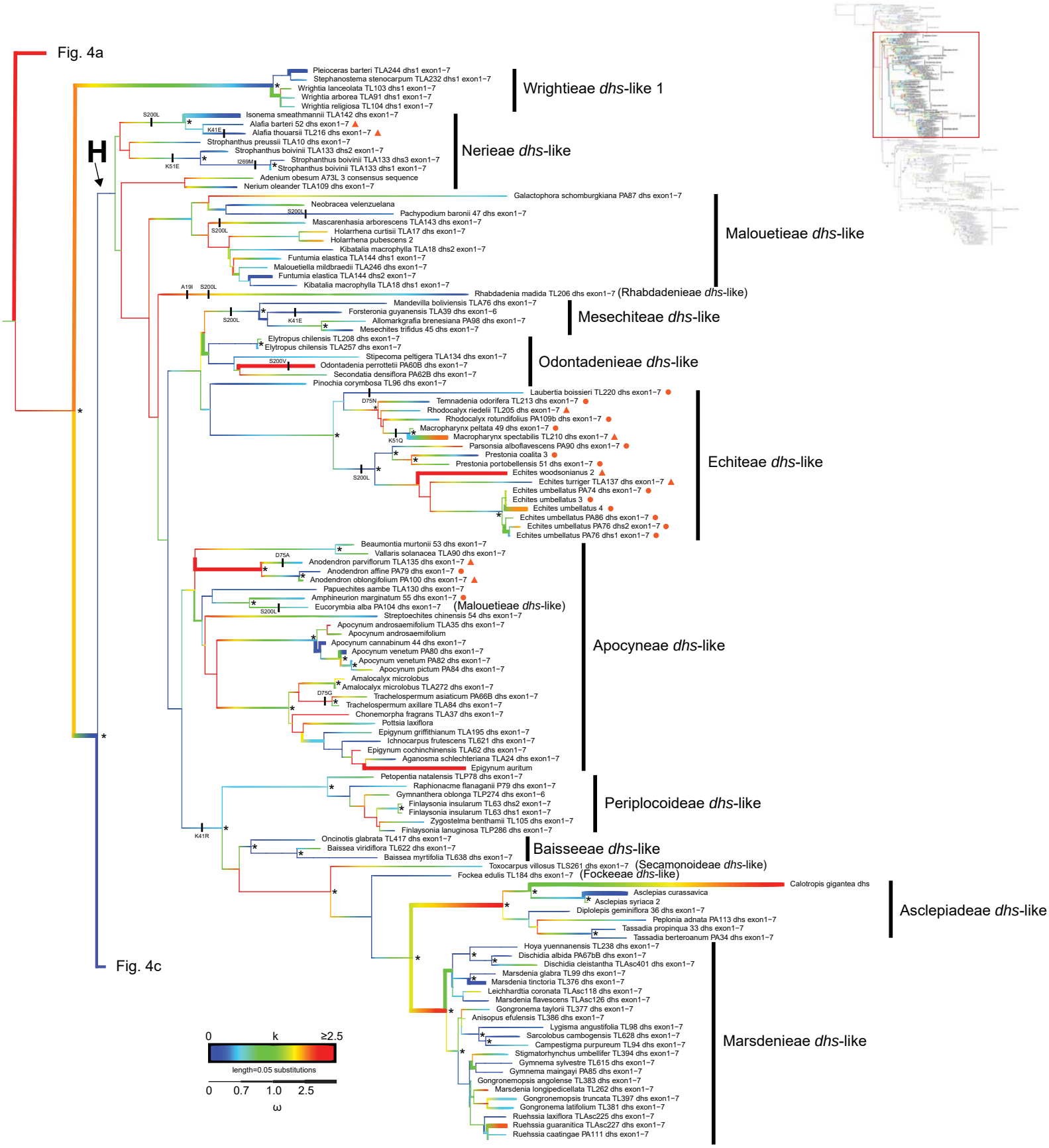
Fig. 4a
Pleioceras barteri TLA244 dhs1 exon1−7
Stephanostema stenocarpum TLA232 dhs1 exon1−7
Wrightia lanceolata TL103 dhs1 exon1−7
Wrightia arborea TLA91 dhs1 exon1−7
Wrightia religiosa TL104 dhs1 exon1−7
Wrightieae dhs-like 1
S200L
Isonema smeathmannii TLA142 dhs exon1−7
Alafia barteri 52 dhs exon1−7
K41E
Alafia thouarsii TL216 dhs exon1−7
Strophanthus preussii TLA10 dhs exon1−7
Strophanthus boivinii TLA133 dhs2 exon1−7
K51E
I269M
Strophanthus boivinii TLA133 dhs3 exon1−7
Strophanthus boivinii TLA133 dhs1 exon1−7
Adenium obesum A73L 3 consensus sequence
Nerium oleander TLA109 dhs exon1−7
Nerieae dhs-like
H
Galactophora schomburgkiana PA87 dhs exon1−7
Neobracea velenzuelana
S200L
Pachypodium baronii 47 dhs exon1−7
Mascarenhasia arborescens TLA143 dhs exon1−7
S200L
Holarrhena curtisii TLA17 dhs exon1−7
Holarrhena pubescens 2
Kibatalia macrophylla TLA18 dhs2 exon1−7
Funtumia elastica TLA144 dhs1 exon1−7
Malouetiella mildbraediii TLA246 dhs exon1−7
Funtumia elastica TLA144 dhs2 exon1−7
Kibatalia macrophylla TLA18 dhs1 exon1−7
Malouetieae dhs-like
A19I
S200L
Rhabdadenia madida TL206 dhs exon1−7
(Rhabdadenieae dhs-like)
Mandevilla boliviensis TLA76 dhs exon1−7
Forsteronia guyanensis TLA39 dhs exon1−6
S200L
K41E
Allomarkgrafia brenesiana PA98 dhs exon1−7
Mesechites trifidus 45 dhs exon1−7
Mesechiteae dhs-like
Elytropus chilensis TL208 dhs exon1−7
Elytropus chilensis TLA257 dhs exon1−7
Stipecoma peltigera TLA134 dhs exon1−7
S200V
Odontadenia perrottetii PA60B dhs exon1−7
Secondatia densiflora PA62B dhs exon1−7
Pinochia corymbosa TL96 dhs exon1−7
Odontadenieae dhs-like
Laubertia boissieri TL220 dhs exon1−7
D75N
Temnadenia odorifera TL213 dhs exon1−7
Rhodocalyx riedelii TL205 dhs exon1−7
Rhodocalyx rotundifolius PA109b dhs exon1−7
Macropharynx peltata 49 dhs exon1−7
K51Q
Macropharynx spectabilis TL210 dhs exon1−7
Parsonsia alboflavescens PA90 dhs exon1−7
Prestonia coalita 3
Prestonia portobellensis 51 dhs exon1−7
S200L
Echites woodsonianus 2
Echites turriger TLA137 dhs exon1−7
Echites umbellatus PA74 dhs exon1−7
Echites umbellatus 3
Echites umbellatus 4
Echites umbellatus PA86 dhs exon1−7
Echites umbellatus PA76 dhs2 exon1−7
Echites umbellatus PA76 dhs1 exon1−7
Echiteae dhs-like
Beaumontia murtonii 53 dhs exon1−7
Vallaris solanacea TLA90 dhs exon1−7
D75A
Anodendron parviflorum TLA135 dhs exon1−7
Anodendron affine PA79 dhs exon1−7
Anodendron oblongifolium PA100 dhs exon1−7
Papuechites aambe TLA130 dhs exon1−7
Amphineurion marginatum 55 dhs exon1−7
Eucorymbia alba PA104 dhs exon1−7
S200L
(Malouetieae dhs-like)
Streptoechites chinensis 54 dhs exon1−7
Apocynum androsaemifolium TLA35 dhs exon1−7
Apocynum androsaemifolium
Apocynum cannabinum 44 dhs exon1−7
Apocynum venetum PA80 dhs exon1−7
Apocynum venetum PA82 dhs exon1−7
Apocynum pictum PA84 dhs exon1−7
Amalocalyx microlobus
Amalocalyx microlobus TLA272 dhs exon1−7
D75G
Trachelospermum asiaticum PA66B dhs exon1−7
Trachelospermum axillare TLA84 dhs exon1−7
Chonemorpha fragrans TLA37 dhs exon1−7
Pottsia laxiflora
Epigynum griffithianum TLA195 dhs exon1−7
Ichnocarpus frutescens TL621 dhs exon1−7
Epigynum cochinchinensis TLA62 dhs exon1−7
Aganosma schlechteriana TLA24 dhs exon1−7
Epigynum auritum
Apocyneae dhs-like
Petopentia natalensis TLP78 dhs exon1−7
Raphionacme flanaganii P79 dhs exon1−7
Gymnanthera oblonga TLP274 dhs exon1−6
Finlaysonia insularum TL63 dhs2 exon1−7
Finlaysonia insularum TL63 dhs1 exon1−7
Zygostelma benthamii TL105 dhs exon1−7
Finlaysonia lanuginosa TLP286 dhs exon1−7
Periplocoideae dhs-like
K41R
Oncinotis glabrata TL417 dhs exon1−7
Baissea viridiflora TL622 dhs exon1−7
Baissea myrtifolia TL638 dhs exon1−7
Baisseeae dhs-like
Toxocarpus villosus TLS261 dhs exon1−7
(Secamonoideae dhs-like)
Fockea edulis TL184 dhs exon1−7
(Fockeeae dhs-like)
Calotropis gigantea dhs
Asclepias curassavica
Asclepias syriaca 2
Diplolepis geminiflora 36 dhs exon1−7
Peplonia adnata PA113 dhs exon1−7
Tassadia propinqua 33 dhs exon1−7
Tassadia berteroanum PA34 dhs exon1−7
Asclepiadeae dhs-like
Hoya yuennanensis TL238 dhs exon1−7
Dischidia albida PA67bB dhs exon1−7
Dischidia cleistantha TLAsc401 dhs exon1−7
Marsdenia glabra TL99 dhs exon1−7
Marsdenia tinctoria TL376 dhs exon1−7
Leichhardtia coronata TLAsc118 dhs exon1−7
Marsdenia flavescens TLAsc126 dhs exon1−7
Gongronema taylorii TL377 dhs exon1−7
Anisopus efulensis TL386 dhs exon1−7
Lygisma angustifolia TL98 dhs exon1−7
Sarcolobus cambogensis TL628 dhs exon1−7
Campestigma purpureum TL94 dhs exon1−7
Stigmatorhynchus umbellifer TL394 dhs exon1−7
Gymnema sylvestre TL615 dhs exon1−7
Gymnema maingayi PA85 dhs exon1−7
Gongronemopsis angolense TL383 dhs exon1−7
Marsdenia longipedicellata TL262 dhs exon1−7
Gongronemopsis truncata TL397 dhs exon1−7
Gongronema latifolium TL381 dhs exon1−7
Ruehssia laxiflora TLAsc225 dhs exon1−7
Ruehssia guaranitica TLAsc227 dhs exon1−7
Ruehssia caatingae PA111 dhs exon1−7
Marsdenieae dhs-like
Fig. 4c
0
k
≥2.5
length=0.05 substitutions
0
0.7
1.0
2.5
ω

Figure 4c

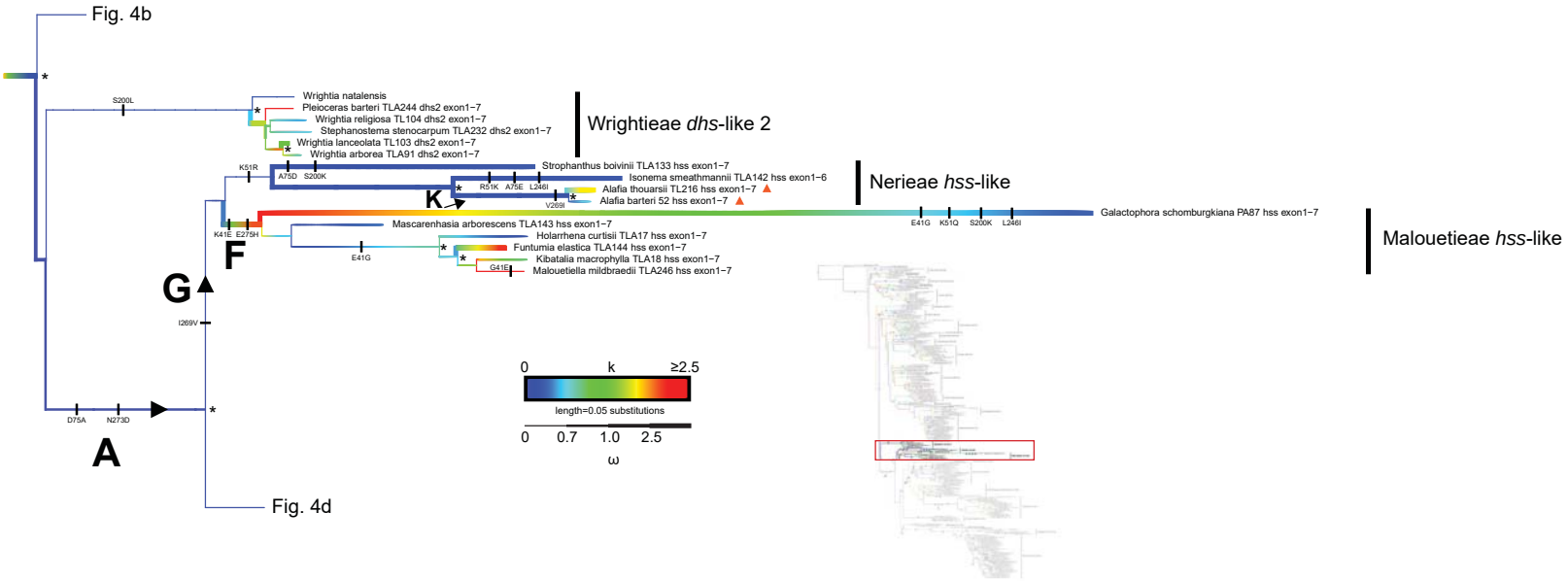
Fig. 4b
S200L
Wrightia natalensis
Pleioceras barteri TLA244 dhs2 exon1−7
Wrightia religiosa TL104 dhs2 exon1−7
Stephanostema stenocarpum TLA232 dhs2 exon1−7
Wrightia lanceolata TL103 dhs2 exon1−7
Wrightia arborea TLA91 dhs2 exon1−7
Wrightieae dhs-like 2
K51R
A75D
S200K
Strophanthus boivinii TLA133 hss exon1−7
Isonema smeathmannii TLA142 hss exon1−6
R51K
A75E
L246I
Alafia thouarsii TL216 hss exon1−7
Alafia barteri 52 hss exon1−7
V269I
K
Nerieae hss-like
K41E
E275H
Mascarenhasia arborescens TLA143 hss exon1−7
Holarrhena curtisii TLA17 hss exon1−7
Funtumia elastica TLA144 hss exon1−7
Kibatalia macrophylla TLA18 hss exon1−7
Malouetiella mildbraedii TLA246 hss exon1−7
E41G
G41E
F
E41G
K51Q
S200K
L246I
Galactophora schomburgkiana PA87 hss exon1−7
Malouetieae hss-like
G
I269V
D75A
N273D
A
Fig. 4d
0
k
≥2.5
length=0.05 substitutions
0
0.7
1.0
2.5
ω

Figure 4d

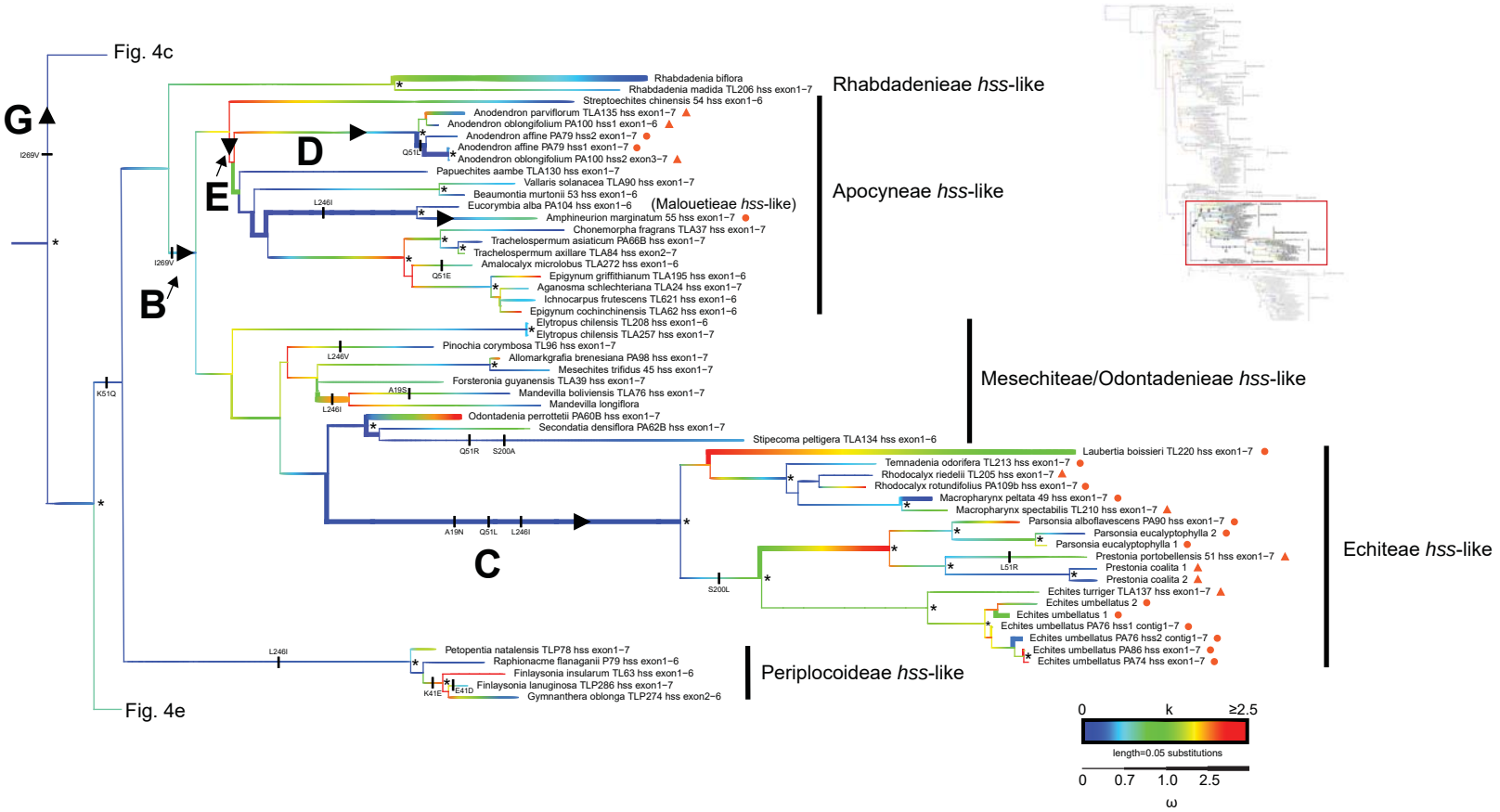

Fig. 4c
G
I269V
D
E
B
C
Rhabdadenia biflora
Rhabdadenia madida TL206 hss exon1−7
Streptoechites chinensis 54 hss exon1−6
Anodendron parviflorum TLA135 hss exon1−7
Anodendron oblongifolium PA100 hss1 exon1−6
Anodendron affine PA79 hss2 exon1−7
Anodendron affine PA79 hss1 exon1−7
Anodendron oblongifolium PA100 hss2 exon3−7
Papuechites aambe TLA130 hss exon1−7
Vallaris solanacea TLA90 hss exon1−7
Beaumontia murtonii 53 hss exon1−6
Eucorymbia alba PA104 hss exon1−6
(Malouetieae hss-like)
Amphineurion marginatum 55 hss exon1−7
Chonemorpha fragrans TLA37 hss exon1−7
Trachelospermum asiaticum PA66B hss exon1−7
Trachelospermum axillare TLA84 hss exon2−7
Amalocalyx microlobus TLA272 hss exon1−6
Epigynum griffithianum TLA195 hss exon1−6
Aganosma schlechteriana TLA24 hss exon1−7
Ichnocarpus frutescens TL621 hss exon1−6
Epigynum cochinchinensis TLA62 hss exon1−6
Elytropus chilensis TL208 hss exon1−6
Elytropus chilensis TLA257 hss exon1−7
Pinochia corymbosa TL96 hss exon1−7
Allomarkgrafia brenesiana PA98 hss exon1−7
Mesechites trifidus 45 hss exon1−7
Forsteronia guyanensis TLA39 hss exon1−7
Mandevilla boliviensis TLA76 hss exon1−7
Mandevilla longiflora
Odontadenia perrottetii PA60B hss exon1−7
Secondatia densiflora PA62B hss exon1−7
Stipecoma peltigera TLA134 hss exon1−6
Laubertia boissieri TL220 hss exon1−7
Temnadenia odorifera TL213 hss exon1−7
Rhodocalyx riedelii TL205 hss exon1−7
Rhodocalyx rotundifolius PA109b hss exon1−7
Macropharynx peltata 49 hss exon1−7
Macropharynx spectabilis TL210 hss exon1−7
Parsonsia alboflavescens PA90 hss exon1−7
Parsonsia eucalyptophylla 2
Parsonsia eucalyptophylla 1
Prestonia portobellensis 51 hss exon1−7
Prestonia coalita 1
Prestonia coalita 2
Echites turriger TLA137 hss exon1−7
Echites umbellatus 2
Echites umbellatus 1
Echites umbellatus PA76 hss1 contig1−7
Echites umbellatus PA76 hss2 contig1−7
Echites umbellatus PA86 hss exon1−7
Echites umbellatus PA74 hss exon1−7
Petopentia natalensis TLP78 hss exon1−7
Raphionacme flanaganii P79 hss exon1−6
Finlaysonia insularum TL63 hss exon1−6
Finlaysonia lanuginosa TLP286 hss exon1−7
Gymnanthera oblonga TLP274 hss exon2−6
Rhabdadenieae hss-like
Apocyneae hss-like
Mesechiteae/Odontadenieae hss-like
Echiteae hss-like
Periplocoideae hss-like
Fig. 4e
0
k
≥2.5
length=0.05 substitutions
0
0.7
1.0
2.5
ω

Figure 4e

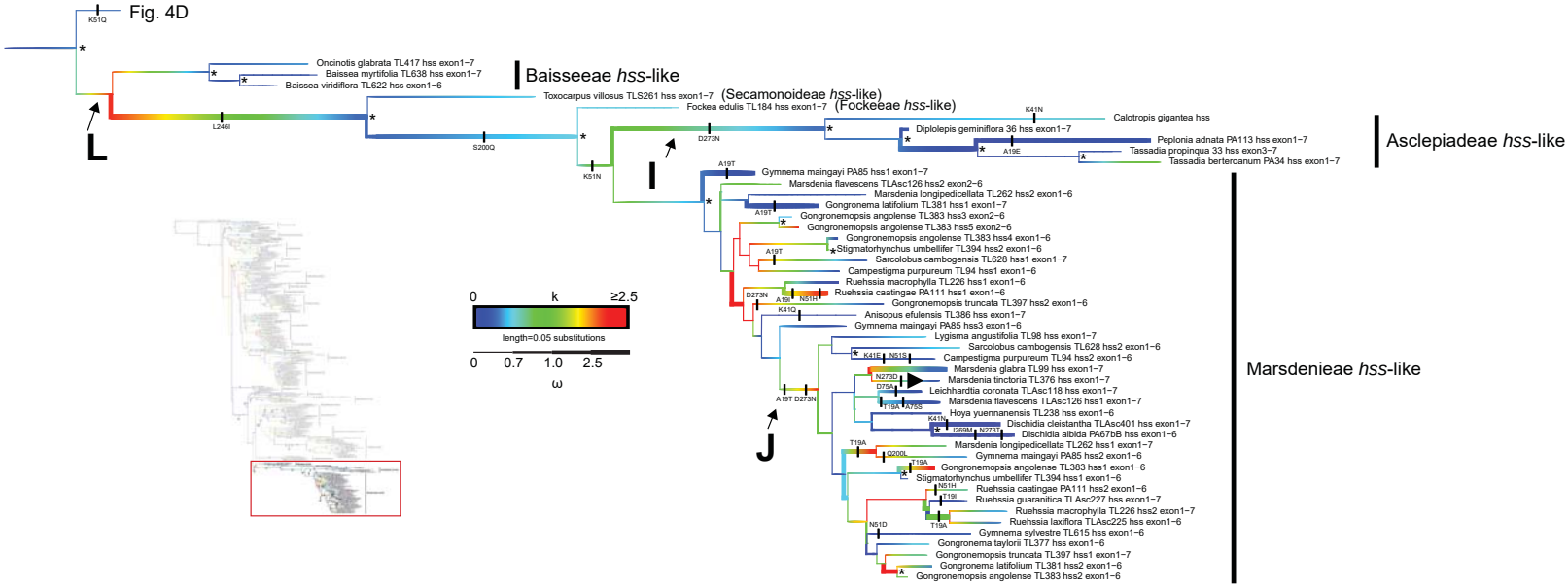
Fig. 4D
K51Q
Oncinotis glabrata TL417 hss exon1−7
Baissea myrtifolia TL638 hss exon1−7
Baissea viridiflora TL622 hss exon1−6
Baisseeae hss-like
Toxocarpus villosus TLS261 hss exon1−7
(Secamonoideae hss-like)
Fockea edulis TL184 hss exon1−7
(Fockeeae hss-like)
K41N
Calotropis gigantea hss
Diplolepis geminiflora 36 hss exon1−7
Peplonia adnata PA113 hss exon1−7
Tassadia propinqua 33 hss exon3−7
Tassadia berteroanum PA34 hss exon1−7
Asclepiadeae hss-like
L
L246I
S200Q
K51N
D273N
A19E
I
A19T
Gymnema maingayi PA85 hss1 exon1−7
Marsdenia flavescens TLAsc126 hss2 exon2−6
Marsdenia longipedicellata TL262 hss2 exon1−6
Gongronema latifolium TL381 hss1 exon1−7
Gongronemopsis angolense TL383 hss3 exon2−6
Gongronemopsis angolense TL383 hss5 exon2−6
Gongronemopsis angolense TL383 hss4 exon1−6
Stigmatorhynchus umbellifer TL394 hss2 exon1−6
Sarcolobus cambogensis TL628 hss1 exon1−7
Campestigma purpureum TL94 hss1 exon1−6
Ruehssia macrophylla TL226 hss1 exon1−6
Ruehssia caatingae PA111 hss1 exon1−6
Gongronemopsis truncata TL397 hss2 exon1−6
Anisopus efulensis TL386 hss exon1−7
Gymnema maingayi PA85 hss3 exon1−6
Lygisma angustifolia TL98 hss exon1−7
Sarcolobus cambogensis TL628 hss2 exon1−6
Campestigma purpureum TL94 hss2 exon1−6
Marsdenia glabra TL99 hss exon1−7
Marsdenia tinctoria TL376 hss exon1−7
Leichhardtia coronata TLAsc118 hss exon1−7
Marsdenia flavescens TLAsc126 hss1 exon1−7
Hoya yuennanensis TL238 hss exon1−6
Dischidia cleistantha TLAsc401 hss exon1−7
Dischidia albida PA67bB hss exon1−6
Marsdenia longipedicellata TL262 hss1 exon1−7
Gymnema maingayi PA85 hss2 exon1−6
Gongronemopsis angolense TL383 hss1 exon1−6
Stigmatorhynchus umbellifer TL394 hss1 exon1−6
Ruehssia caatingae PA111 hss2 exon1−6
Ruehssia guaranitica TLAsc227 hss exon1−7
Ruehssia macrophylla TL226 hss2 exon1−7
Ruehssia laxiflora TLAsc225 hss exon1−6
Gymnema sylvestre TL615 hss exon1−6
Gongronema taylorii TL377 hss exon1−6
Gongronemopsis truncata TL397 hss1 exon1−7
Gongronema latifolium TL381 hss2 exon1−6
Gongronemopsis angolense TL383 hss2 exon1−6
Marsdenieae hss-like
J
0
k
≥2.5
length=0.05 substitutions
0
0.7
1.0
2.5
ω