Efficient nonenzymatic cyclization and domain shuffling drive pyrrolopyrazine diversity from truncated variants of a fungal NRPS

Daniel Berry[a,b], Wade Mace[c], Katrin Grage[a], Frank Wesche[d], Sagar Gore[e], Christopher L. Schardl[f], Carolyn A. Young[g], Paul P. Dijkwel[a,b], Adrian Leuchtmann[h], Helge B. Bode[d,i,j,1], and Barry Scott[a,b,1]

[a]School of Fundamental Sciences, Massey University, Palmerston North 4442, New Zealand; [b]Bioprotection Research Centre, Massey University, Palmerston North 4442, New Zealand; [c]Grasslands Research Centre, AgResearch Ltd., Palmerston North 4442, New Zealand; [d]Fachbereich Biowissenschaften, Molekulare Biotechnologie, Goethe Universität Frankfurt, 60438 Frankfurt am Main, Germany; [e]Leibniz Institute for Natural Product Research and Infection Biology, Hans Knöll Institute, 07745 Jena, Germany; [f]Department of Plant Pathology, University of Kentucky, Lexington, KY 40506; [g]Noble Research Institute, LLC, Ardmore, OK 73401; [h]Institute of Integrative Biology, ETH Zürich, CH-8092 Zürich, Switzerland; [i]Buchmann Institute for Molecular Life Sciences, Goethe-Universität, 60438 Frankfurt am Main, Germany; and [j]Landes-Offensive zur Entwicklung Wissenschaftlich-Ökonomischer Exzellenz (LOEWE) Centre for Translational Biodiversity Genomics, 60325 Frankfurt am Main, Germany

Edited by Jay C. Dunlap, Geisel School of Medicine at Dartmouth, Hanover, NH, and approved November 11, 2019 (received for review August 7, 2019)

Nonribosomal peptide synthetases (NRPSs) generate the core peptide scaffolds of many natural products. These include small cyclic dipeptides such as the insect feeding deterrent peramine, which is a pyrrolopyrazine (PPZ) produced by grass-endophytic *Epichloë* fungi. Biosynthesis of peramine is catalyzed by the 2-module NRPS, PpzA-1, which has a C-terminal reductase (R) domain that is required for reductive release and cyclization of the NRPS-tethered dipeptidyl-thioester intermediate. However, some PpzA variants lack this R domain due to insertion of a transposable element into the 3′ end of *ppzA*. We demonstrate here that these truncated PpzA variants utilize nonenzymatic cyclization of the dipeptidyl thioester to a 2,5-diketopiperazine (DKP) to synthesize a range of novel PPZ products. Truncation of the R domain is sufficient to subfunctionalize PpzA-1 into a dedicated DKP synthetase, exemplified by the truncated variant, PpzA-2, which has also evolved altered substrate specificity and reduced *N*-methyltransferase activity relative to PpzA-1. Further allelic diversity has been generated by recombination-mediated domain shuffling between *ppzA-1* and *ppzA-2*, resulting in the *ppzA-3* and *ppzA-4* alleles, each of which encodes synthesis of a unique PPZ metabolite. This research establishes that efficient NRPS-catalyzed DKP biosynthesis can occur in vivo through nonenzymatic dipeptidyl cyclization and presents a remarkably clean example of NRPS evolution through recombinant exchange of functionally divergent domains. This work highlights that allelic variants of a single NRPS can result in a surprising level of secondary metabolite diversity comparable to that observed for some gene clusters.

nonribosomal peptide synthetase | secondary metabolism | diketopiperazine | pyrrolopyrazine | allelic neofunctionalization

Nonribosomal peptide synthetases (NRPSs) are large multimodular proteins that produce a huge variety of bioactive nonribosomal peptide (NRP) natural products, including small cyclic dipeptides such as the 2,5-diketopiperazines (1) and the pyrazinones (2–4). NRP biosynthesis occurs via a thiotemplate mechanism, with each module responsible for the colinear incorporation of a single amino acid substrate into the growing peptide chain (5–7). Each module contains an adenylylation (A) domain, which binds and activates a specific amino acid substrate using adenosine 5′-triphosphate, and a thiolation (T) domain with a 4′-phosphopantetheinyl cofactor to which the activated aminoacyl substrate is tethered via a thioester bond. Modules downstream of the N-terminal "initiation" module also begin with a condensation (C) domain, which catalyzes peptide bond formation between adjacent thiotethered aminoacyl or peptidyl intermediates, and all modules may optionally contain one or more accessorizing domains that modify the aminoacyl intermediate, such as the *N*-methyltransferase (M) domain. Most NRPSs end with a termination domain, such as a thioesterase or reductase (R) domain, which catalyzes release of the mature peptidyl chain (8).

Peramine (**1a**; Fig. 1) is a pyrrolopyrazine-1-one metabolite (PPZ-1-one) synthesized by fungi of genus *Epichloë* (9), which are symbiotic endophytes of cool-season grasses (10). *Epichloë* spp. produce a range of bioactive secondary metabolites that protect their grass hosts against herbivory (10), and production of **1a** is agriculturally desirable due to the protection it provides against the pastoral pest insect *Listronotus bonariensis* (Argentine stem weevil) (9, 11, 12). Biosynthesis of **1a** is catalyzed by the 2-module NRPS "PpzA-1," which has the domain structure A_1-T_1-C_2-A_2-M_2-T_2-R_2, where numbers indicate the module to which each domain belongs (2). PpzA-1 also has a partial C domain at its N terminus that is thought to be an evolutionary relic required

Significance

Nonribosomal peptide synthetases (NRPSs) synthesize the core peptide scaffold of many natural products. These include small cyclic dipeptides such as peramine, which is a potent insect feeding deterrent synthesized by the 2-module NRPS PpzA-1 from grass endophytic fungi. Here we identify several new PpzA variants lacking the C-terminal product release domain of PpzA-1 that instead utilize efficient nonenzymatic cyclization of the NRPS-tethered dipeptidyl-thioester intermediate to release a range of diketopiperazine-containing products. The metabolic diversity generated from these truncated variants is the result of altered biosynthetic activities combined with recombination-mediated domain shuffling. This work highlights that allelic variants of a single NRPS can result in a surprising level of secondary metabolite diversity comparable to that observed for some gene clusters.

Author contributions: D.B., S.G., C.L.S., C.A.Y., P.P.D., H.B.B., and B.S. designed research; D.B., W.M., K.G., F.W., S.G., C.L.S., and H.B.B. performed research; A.L. contributed new reagents/analytic tools; D.B., W.M., K.G., H.B.B., and B.S. analyzed data; and D.B. and B.S. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Data deposition: DNA sequence data reported in this paper have been deposited in the GenBank database, https://www.ncbi.nlm.nih.gov/genbank/ (accession nos. MN605951–MN605962).

[1]To whom correspondence may be addressed. Email: h.bode@bio.uni-frankfurt.de or d.b.scott@massey.ac.nz.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1913080116/-/DCSupplemental.

First published December 4, 2019.

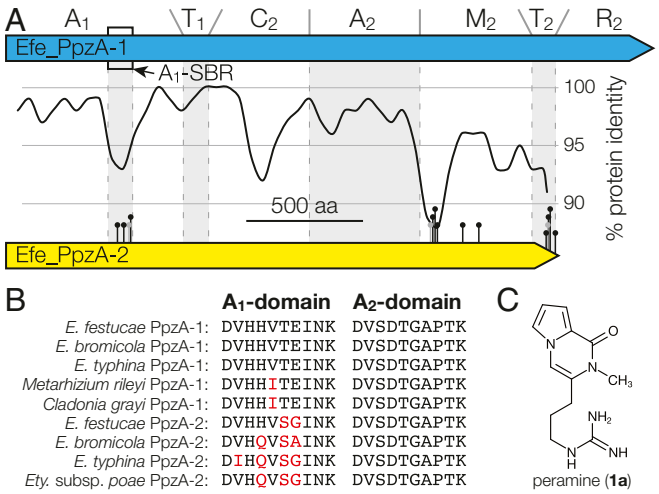

Fig. 1. Functionally divergent regions of PpzA-1 and PpzA-2 proteins. (*A*) Sliding-window analysis (width = 100, step = 50) of protein sequence conservation across an alignment of Efe_PpzA-1 and Efe_PpzA-2 protein sequences. The Efe_PpzA-1 protein map is annotated to show approximate domain boundaries and the A_1-domain substrate binding region (A_1-SBR). The positions of substitutions that are both conserved among PpzA-2 proteins and unique to PpzA-2 proteins are annotated on the Efe_PpzA-2 protein map with black pins. Gray pins indicate residues which are always substituted in PpzA-2 proteins relative to PpzA-1 proteins but whose identities are not fully conserved among PpzA-2 proteins. (*B*) A_1- and A_2-domain NRPS code residues for representative PpzA-1 and PpzA-2 protein sequences. Residues that differ from *Epichloë* spp. PpzA-1 codes are highlighted in red. (*C*) Structure of peramine (**1a**).

to maintain A-domain stability (13). The first module of PpzA-1 is thought to incorporate a pyrrolidine-containing amino acid substrate, while the second module incorporates and *N*-methylates arginine (Arg). A peptide bond is then formed between these aminoacyl substrates, and the resulting dipeptidyl thioester is reductively released by the R_2 domain as a dipeptide aldehyde intermediate that is then thought to undergo spontaneous cyclization, rearrangement, and oxidation reactions to form **1a**.

This study focuses on allelic variants of *ppzA* that contain a transposable element insertion into the 3′ end of the gene, resulting in deletion of all sequence encoding the C-terminal R_2 domain (14–16). While initially assumed to be pseudogenes, these truncated "*ppzA-2*" alleles were subsequently found to be widely distributed across multiple *Epichloë* spp. (14) and are still expressed at levels comparable to *ppzA-1* (17, 18). The *ppzA-1* and *ppzA-2* alleles also exhibit transspecies polymorphisms (TSP), suggesting that balancing selection has maintained both alleles since the emergence of *ppzA-2* in a common ancestor of most *Epichloë* spp. (14). We therefore hypothesized that *ppzA-2* alleles may encode functional NRPSs with novel biosynthetic activities. As bioprotective *Epichloë* secondary metabolism genes such as *ppzA* are typically only expressed by *Epichloë* spp. *in planta* (2, 19), we use a heterologous expression system to identify the PPZ metabolites produced by different PpzA variants. We also investigate the biosynthetic mechanisms underpinning PPZ biosynthesis and demonstrate how recombination between different *ppzA* alleles has driven PPZ biosynthetic diversity. Note that while the peramine synthetase-encoding gene was originally abbreviated as "*perA*" (2), we suggest a modification to "*ppzA*" to better reflect the class of metabolites produced, with allele *ppzA-1* being synonymous with *perA*.

Results

Identifying Divergent Regions in PpzA-2 Proteins. Although PpzA-2 proteins lack the C-terminal R_2 domain found in PpzA-1, the domain structures of PpzA-1 and PpzA-2 proteins are otherwise identical. Sliding-window analysis showed that conservation between *Epichloë festucae* PpzA-1 (Efe_PpzA-1) and PpzA-2 (Efe_PpzA-2) proteins approaches 100% sequence identity in many regions (Fig. 1*A*). However, discrete regions within the A_1, C_2, M_2, and T_2 domains exhibit increased divergence, and substitutions that are both conserved among and unique to PpzA-2 proteins (PpzA-2–conserved substitutions) are exclusively located within these divergent regions (Fig. 1*A* and *SI Appendix*, Figs. S1–S4). Efe_PpzA-2 sequence divergence within the A_1 domain primarily affected the substrate-binding region (SBR), which is located between conserved A-domain motifs A4 and A5 (*SI Appendix*, Fig. S1) (20). This SBR contains 9 of the 10 "NRPS code" residues, which are the primary determinants of A-domain substrate specificity (21). The NRPS code residues of PpzA-1 A_1 domains are absolutely conserved across all *Epichloë* spp.; however, 2 of these residues are always substituted in PpzA-2 proteins, and additional substitutions are observed in PpzA-2 proteins from *Epichloë bromicola* and *Epichloë typhina* strains (Fig. 1*B* and *SI Appendix*, Fig. S1). This lack of NRPS code conservation suggests that the A_1 domains from PpzA-1 and PpzA-2 proteins may recognize different amino acid substrates.

Recombinational Shuffling Has Generated ppzA Allelic Diversity. Sliding-window phylogenetic comparison between different *ppzA* alleles indicated that the *Efe_ppzA-2* is actually a chimeric allele resulting from several recombinational cross-over events between ancestral *ppzA-1* and *ppzA-2* sequences (Fig. 2*A*). Approximately 25% of *Efe_ppzA-2*, including the A_1 SBR, the 5′ end of the M_2 domain, the T_2 domain, and a small part of the A_2-encoding region, appears to be inherited from an ancestral *ppzA-2* allele. The remaining 75% of *Efe_ppzA-2*, including the T_1 and C_2 domains, the majority of the A_1 and A_2 domains, and the 3′ end of the M_2 domain, appears to descend from an ancestral *ppzA-1* allele (Fig. 2*A*). Very similar patterns of recombination were observed in *ppzA-2* alleles from *E. bromicola* (*Ebr_ppzA-2*); however, the *ppzA-1*–derived regions of *Ebr_ppzA-2* and *Efe_ppzA-2* alleles still grouped with extant *Ebr_ppzA-1* and *Efe_ppzA-1* sequences, respectively (*SI Appendix*, Fig. S5). This phylogeny therefore supports a convergent evolution model where *Ebr_ppzA-2* and *Efe_ppzA-2* were generated by independent yet equivalent recombination events that occurred after divergence of the *E. bromicola* and *E. festucae* lineages. No evidence of widespread recombination between *ppzA* alleles was observed within the *E. typhina* clade (*SI Appendix*, Fig. S5).

Some *ppzA* alleles from *Epichloë baconii*, *E. bromicola*, and *E. festucae* isolates contained the same 3′ deletion found in *ppzA-2* alleles yet exhibited recombination patterns that were different from their *ppzA-2* counterparts and were therefore defined as "*ppzA-3*" and "*ppzA-4*" alleles (Fig. 2). The *ppzA-3* allele was identified from some *E. bromicola* and *E. baconii* strains and differs from its *ppzA-2* counterparts in that it retains a *ppzA-1*–type A_1 SBR-encoding region (Fig. 2*A*). Phylogenetic sequence analysis suggests that these *Ebr_ppzA-3* and *Eba_ppzA-3* alleles represent intermediates during the evolution of *Ebr_ppzA-2* and *Efe_ppzA-2*, respectively (Fig. 2*B* and *SI Appendix*, Fig. S5); however, both alleles have also been retained within their respective host populations. The *ppzA-4* allele was only identified in *Epichloë siegelii*, which is an asexual allopolyploid hybrid species of *E. festucae* and *E. bromicola* (22). This *Esi_ppzA-4* allele is from the *E. festucae*-derived portion of the *E. siegelii* genome and appears to be an *Efe_ppzA-2* sequence that underwent an additional recombination event that replaced most of the M_2-encoding region with sequence from an *Efe_ppzA-1* donor (Fig. 2 *A* and *B* and *SI Appendix*, Fig. S5). Because *E. siegelii* is asexual and lacks a suitable *Efe_ppzA-1* donor sequence, this additional recombination event likely predates the emergence of

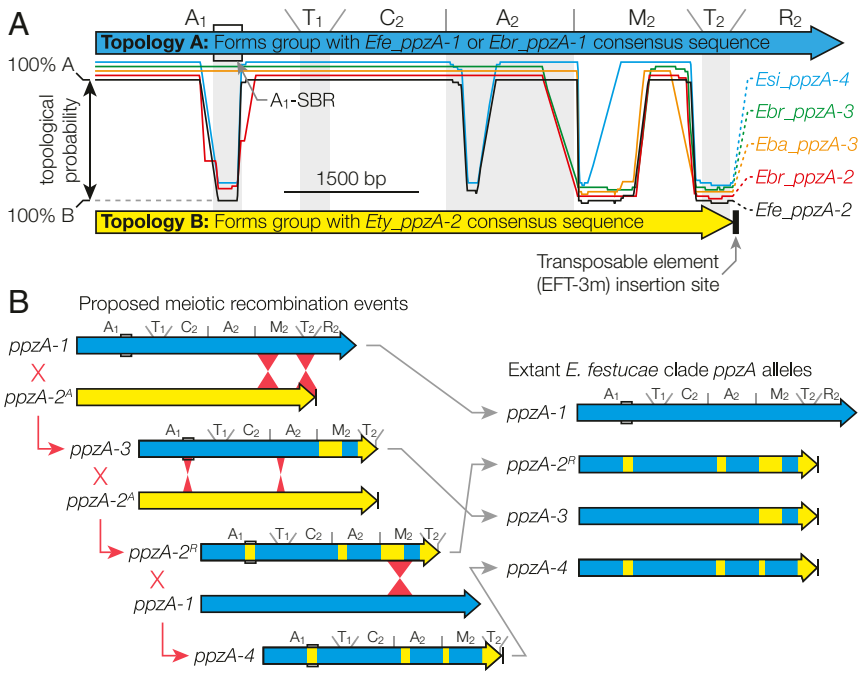

Fig. 2. Recombination between *ppzA* alleles generated additional diversity. (*A*) Sliding-window comparison of listed alleles to consensus sequences for *Ety_ppzA-1*, *Ety_ppzA-2*, *Efe_ppzA-1* (for *Efe_ppzA_2*, *Eba_ppzA-3*, and *Esi_ppzA-4*), or *Ebr_ppzA-1* (for *Ebr_ppzA-2* and *Ebr_ppzA-3*) using DualBrothers 1.1.5 (52) to identify spatial variations in phylogenetic topology (200-bp window, 50-bp step). Graph lines are slightly offset on the vertical axis to aid visual clarity. (*B*) Proposed order of ancestral meiotic recombination events between *ppzA* alleles in the *E. festucae* clade lineage, where *E. baconii* is defined as an *E. festucae* clade member. Ancestral and recombination-derived *ppzA-2* alleles are differentiated as *ppzA-2*^{A} and *ppzA-2*^{R}, respectively.

E. siegelii, although an equivalent *ppzA-4* allele has not been observed in any *E. festucae* isolates to date.

The Truncated Allele *ppzA-2* Encodes a Pyrrolopyrazine-1,4-Dione Synthetase. The fungus *Penicillium paxilli* strain PN2013 was utilized as a naive heterologous expression host to identify the metabolite products of the enzymes encoded by different *ppzA* alleles, with **1a**-producing PN2013/*Efe_ppzA-1* control strains obtained from a previous study (23). As expected, PN2013 transformants expressing *Efe_ppzA-2* from *E. festucae* strain E189 or *Ety_ppzA-2* from *E. typhina* subsp. *poae* strain E1022 did not produce **1a** (*SI Appendix*, Table S1); however, metabolomic analysis identified a 254.16 *m/z* metabolite exclusively produced by these *ppzA-2*–expressing strains (*SI Appendix*, *Supplementary Note*). Liquid chromatography-coupled tandem mass spectrometry comparison to a synthetic standard and NMR analysis of the purified natural product determined that this metabolite was the PPZ-1,4-dione molecule cyclo(L-Pro, L-Arg) (**2a**) (Fig. 3 *A* and *B* and *SI Appendix*, *Supplementary Note*). Analysis of extracts from grass material confirmed that **2a** was present in all samples infected with *ppzA-2*-genotype *Epichloë* isolates, whereas **2a** was not observed in any samples infected with *ppzA-1*-genotype *Epichloë* isolates or in uninfected plants (Table 1). The *Efe_ppzA-2* gene was deleted in *E. festucae* strain E189 (*SI Appendix*, Fig. S6), and loss of **2a** was confirmed in grass samples infected with these Δ*Efe_ppzA-2* mutants (Fig. 3*C* and *SI Appendix*, Table S2).

Mean production of **2a** was 10-fold higher in PN2013/*Ety_ppzA-2* cultures compared to PN2013/*Efe_ppzA-2* cultures ($P = 0.038$; Fig. 3*D*), and plant material infected with *Ety_ppzA-2*-genotype endophytes also exhibited the highest **2a** concentrations (Table 1 and *SI Appendix*, Table S3), suggesting that Ety_PpzA-2 may be a more efficient **2a** synthetase. The absence of an Arg *N*-methyl group in the structure of **2a** (Fig. 3*A*) suggested that the PpzA-2–specific M_2-domain substitutions may have eliminated *N*-methyltransferase activity (Fig. 1*A* and *SI Appendix*, Fig. S3). Targeted analysis revealed that the *N*-methylated PPZ-1,4-dione cyclo(Pro, meArg) (**2b**; Fig. 3*A*) was also produced by *Efe_ppzA-2* strains, but not by *Ety_ppzA-2* strains (Fig. 3*D* and *SI Appendix*, *Supplementary Note*). Synthetic allele *ppzA-S1* was generated by replacing the M_2-encoding sequence of *Efe_ppzA-2* with its *Efe_ppzA-1* equivalent to investigate whether the production of **2a** by Efe_PpzA-2 was due to reduced *N*-methyltransferase activity. This abolished production of **2a** and massively increased production of **2b** (Fig. 3*D*), demonstrating that *N*-methyltransferase activity has been reduced in Efe_PpzA-2 by substitutions within the M_2 domain. Additional lineage-specific substitutions may explain the complete loss of M_2-domain function in Ety_PpzA-2 (*SI Appendix*, Fig. S3).

PpzA-1 and PpzA-2 Proteins Bind Different A_1-Domain Substrates. Structural conservation between **1a** and **2a-b** (Fig. 3*A*) suggests that the A_2 domains of PpzA-1 and PpzA-2 proteins share L-Arg as substrate, and this is supported by the absolute conservation of the substrate-defining NRPS code between A_2 domains from all PpzA proteins (Fig. 1*B*). The PpzA-2 A_1-domain substrate can similarly be inferred as L-Pro from the structures of **2a-b**; however, the spontaneous pyrrolidine oxidation step proposed during **1a** biosynthesis (2) and different NRPS codes exhibited by A_1 domains from PpzA-1 vs. PpzA-2 proteins (Fig. 1*B*) mean that the PpzA-1 A_1-domain substrate cannot be determined with certainty. Identification of the cognate PpzA-1 A_1-domain substrate was therefore attempted by feeding several candidate amino acids to PN2013/*Efe_ppzA-1* cultures to assess their effects on **1a** production. Of these, only *trans*-4-hydroxy-L-proline (T4HP) and *cis*-4-hydroxy-D-proline (C4HP) had a substantial impact, with each of these 4-hydroxy-proline (4HP) stereoisomers increasing **1a** production 50-fold relative to water-fed controls (Table 2). Modeling the PpzA-1 A_1-domain structure places the NRPS code residues Thr500 and Glu521 in a position that would enable hydrogen bonding interactions with the pyrrolidine-hydroxy

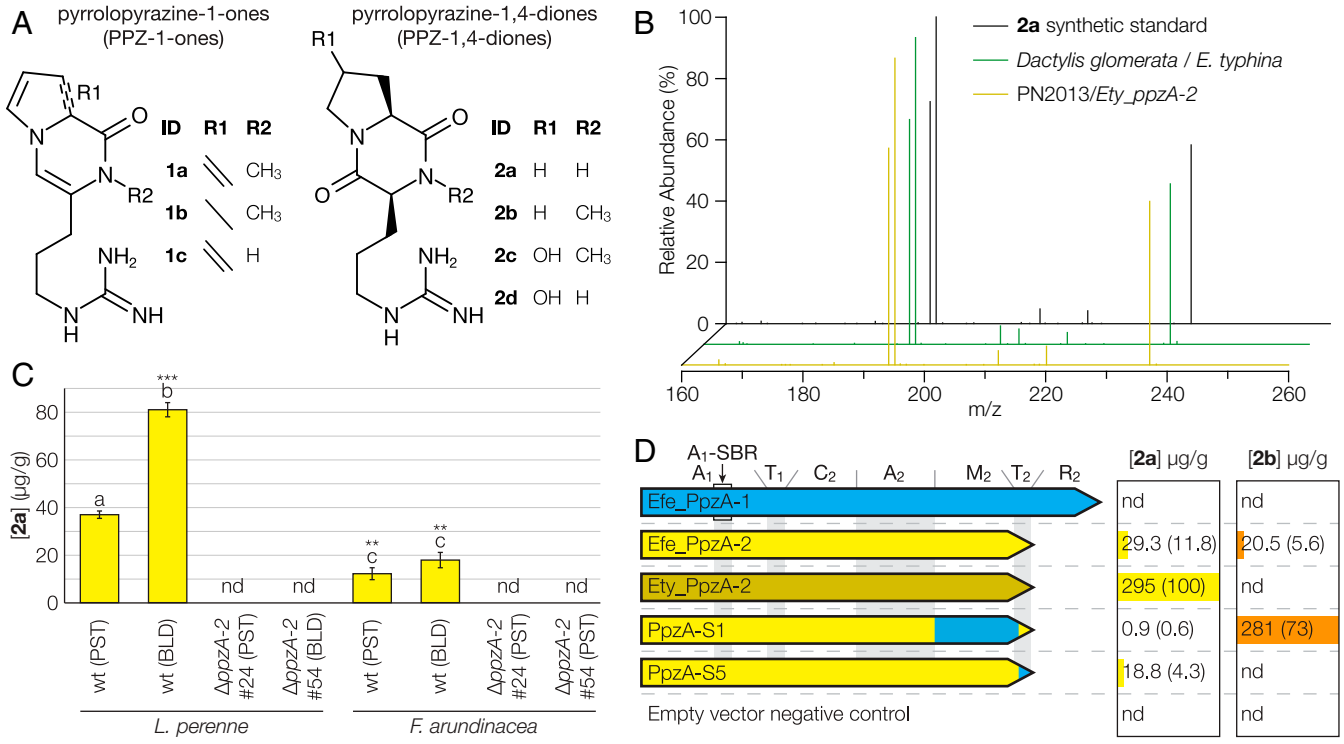

Fig. 3. Identification and characterization of PPZ-1,4-diones synthesized by PpzA-2 proteins. (*A*) Polymorphic structures showing all PPZ metabolites described in this study. (*B*) Comparison of MS^2 spectra generated at 35% normalized collision energy for a synthetic **2a** standard vs. a 254.16 *m/z* metabolite extracted from *E. typhina*-infected plant material and PN2013/*Ety_ppzA-2* mycelia. (*C*) Concentration of **2a** in blade (BLD) or pseudostem (PST) tissue from 2 different grass hosts infected with wild-type *E. festucae* E189 or 1 of 2 Δ*ppzA-2* mutants. Error bars show the SEM; "nd" indicates where **2a** was not detected in a sample (limit of detection 0.05 μg/g). (*D*) Concentration of metabolites **2a** and **2b** in extracts from cultures of *P. paxilli* strains that express the PpzA proteins shown to the left of the image. Protein maps are annotated with domain boundaries, and synthetic hybrid proteins are colored to indicate the natural PpzA protein from which each region is derived. Concentrations are averaged across at least 3 independent transformants for each *ppzA* expression construct, with the SEM shown in parentheses. Metabolites that were not detected are annotated "nd" (limit of detection 0.05 μg/g). Colored bars illustrate the relative between-sample concentration for each metabolite. Concentrations for **2b** are estimates based on the response factor of synthetic **2a**. $^{**}P < 0.01$ and $^{***}P < 0.001$.

group of T4HP and/or C4HP (*SI Appendix*, Fig. S7). In contrast, Thr500 is substituted for Ser and Glu521 is substituted for Gly/Ala in PpzA-2 proteins, presumably altering the binding pocket shape and charged microenvironment to favor Pro as substrate. Additional supplementation of PN2013/*Efe_ppzA-1* cultures with L-Arg did not further increase **1a** production over 4HP feeding alone (Table 2),

Table 1. PPZ profile of grasses infected with representative *Epichloë* strains*

				PPZ concentration†, μg/g						
Allele‡	Endophyte	Strain	Host	[1a]	[1b]§	[1c]§	[2a]	[2b]¶	[2c]¶	[2d]¶
ppzA-1	*E. bromicola*	NFe1	*Hordeum bogdanii*	26.4	0.2	nd	nd	nd	1.0	0.4
ppzA-1	*E. festucae* var. *lolii*	AR5	*Lolium perenne*	19.1	nd	nd	nd	nd	6.0	nd
ppzA-1	*E. typhina*	E8	*Lolium perenne*	347	0.3	LOQ	4.3	4.0	58.3	nd
ppzA-1	*E. typhina* ssp. *poae*	NFe76	*Bromus laevipes*	101	LOQ	nd	nd	nd	18.5	nd
ppzA-2	*E. festucae*	E189	*Festuca rubra* ssp. *rubra*	nd	nd	nd	19.3	1.2	nd	nd
ppzA-2	*E. typhina*	AL1218	*Dactylis glomerata*	nd	nd	nd	699	nd	nd	0.6
ppzA-2	*E. typhina* ssp. *poae*	AL9921/1	*Poa nemoralis*	nd	nd	nd	32.9	nd	nd	nd
ppzA-3	*E. baconii*	E424	*Agrostis tenuis*	nd	nd	nd	nd	nd	nd	0.4
ppzA-3	*E. bromicola*	NFe7	*Hordeum brevisbulatum*	nd	nd	nd	nd	nd	nd	9.6
ppzA-4	*E. siegelii*	e915	*Festuca arundinacea*	nd	nd	nd	nd	515	0.7	nd
ppzA-5	*E. uncinata*	e167	*Festuca pratensis*	nd	nd	nd	nd	nd	2.5	nd
N/A	uninfected	N/A	*Bromus laevipes*	nd	nd	nd	nd	nd	nd	nd

*Results from additional associations and some replicates are in *SI Appendix*, Table S2.
†Metabolites that were not detected are annotated "nd" (limit of detection 0.05 μg/g). LOQ indicates detection of a metabolite at a concentration below the limit of quantitation (0.2 μg/g).
‡As determined by sequence analysis.
§Approximate concentration based on response factor of synthetic **1a**.
¶Approximate concentration based on response factor of synthetic **2a**.

Table 2. Substrate feeding effects on PPZ production

Allele	Medium[†]	PPZ concentration*, μg/g						
		[1a]	[1b][‡]	[1c][‡]	[2a]	[2b][§]	[2c][§]	[2d][§]
Efe_ppzA-1	CD + H_2O	0.3	nd	nd	nd	nd	nd	nd
Efe_ppzA-1	CD + L-Glu	LOQ	nd	nd	nd	nd	nd	nd
Efe_ppzA-1	CD + L-Pro	1.0	nd	nd	nd	2.4	nd	nd
Efe_ppzA-1	CD + P2C	0.5	nd	nd	nd	nd	nd	nd
Efe_ppzA-1	CD + C4HP	21.1	LOQ	nd	nd	nd	3.6	nd
Efe_ppzA-1	CD + T4HP	20.8	LOQ	nd	nd	nd	4.7	nd
Efe_ppzA-1	CD + T4HP, L-Arg	15.3	LOQ	nd	nd	nd	3.4	nd
Efe_ppzA-2	CD + H_2O	NT	NT	NT	28.6	8.0	nd	nd
Efe_ppzA-2	CD + L-Pro	NT	NT	NT	23.0	6.9	nd	nd
Efe_ppzA-2	CD + P2C	NT	NT	NT	26.5	8.3	nd	nd
Efe_ppzA-2	CD + T4HP	NT	NT	NT	18.3	3.4	1.2	3.4
neg. ctrl	CD + L-Pro	nd	nd	nd	nd	nd	nd	nd
neg. ctrl	CD + T4HP	nd	nd	nd	nd	nd	nd	nd

*PPZ concentrations are averaged across cultures of 3 independent *P. paxilli* strains transformed with *Efe_ppzA-1*, *Efe_ppzA-2*, or empty pRS426 vector (negative control). Metabolites that were not detected are annotated "nd" (limit of detection 0.05 μg/g). LOQ indicates detection of a metabolite at a concentration below the limit of quantitation (0.2 μg/g). NT, not tested.

[†]Cultures were grown in 50 mL Czapek Dox liquid medium under standard conditions with feeding of 2.4×10^{-4} mol each substrate after 4 and 5 d growth.

[‡]Approximate concentration based on response factor of synthetic **1a**.

[§]Approximate concentration based on response factor of synthetic **2a**.

indicating that L-Arg availability was not limiting **1a** production. T4HP feeding was therefore used in subsequent experiments to improve production of 4HP-derived metabolites.

PN2013/*Efe_ppzA-1* cultures fed with 4HP also produced the PPZ-1,4-dione cyclo(4HP, meArg) (**2c**; Fig. 3*A* and *SI Appendix*, *Supplementary Note*), which contains a hydroxy group attached to the pyrrolidine ring. Although PpzA-1 had previously only been reported to synthesize **1a** (2), **2c** was subsequently identified in all *ppzA-1*-genotype samples with sufficiently high **1a** concentrations. The hydrogenated PPZ-1-one **1b** (Fig. 3*A* and *SI Appendix*, *Supplementary Note*), which is likely an intermediate that exists in weak equilibrium with **1a**, was also detected in 4HP-fed cultures (Table 2). Additionally, feeding PN2013/*Efe_ppzA-1* cultures with L-Pro induced production of the PPZ-1,4-dione **2b** (Table 2), indicating that the Efe_PpzA-1 A_1 domain exhibits some weak specificity toward this substrate. Feeding PN2013/*Efe_ppzA-2* cultures with T4HP similarly induced production of the hydroxylated PPZ-1,4-diones **2c** and cyclo(4HP, Pro) (**2d**; Fig. 3*A* and *SI Appendix*, *Supplementary Note*), indicating the Efe_PpzA-2 A_1 domain retains residual specificity toward 4HP (Table 2). Unlike production of **1a** by PN2013/*Efe_ppzA-1* cultures, PN2013/*Efe_ppzA-2* cultures did not appear to be substrate-limited for production of **2a** and **2b**, as feeding L-Pro and L-Arg did not dramatically increase **2a** concentrations (41.3 μg/g with L-Pro/L-Arg feeding vs. 26.5 μg/g water-fed control). Collectively, these results show that 4HP is likely the substrate of the PpzA-1 A_1 domain, and that PpzA-1 can synthesize both PPZ-1-one and PPZ-1,4-dione products.

Truncated Alleles *ppzA-3* and *ppzA-4* also Encode Functionally Distinct PPZ-1,4-Dione Synthetases. As the A_1 domains of PpzA-3 proteins retain PpzA-1–like NRPS codes (*SI Appendix*, Fig. S1), these were predicted to bind 4HP substrate. Expression of *Eba_ppzA-3* from *E. baconii* strain As6 in *P. paxilli* induced production of the hydroxylated PPZ-1,4-diones **2d** and **2c** (Fig. 4). Replacement of the A_1 SBR-encoding region of *Efe_ppzA-2* with the equivalent sequence from *Efe_ppzA-1* was used to generate *ppzA-S2*, which is a synthetic analog of *ppzA-3*. Expression of *ppzA-S2* also induced production of **2c** and **2d** in *P. paxilli* (Fig. 4). Analysis of *Epichloë*-infected plant material also showed that **2d** was exclusive to samples infected with *ppzA-3*-genotype *E. baconii* or *E. bromicola* strains (Table 1), although concentrations were much higher in the *E. bromicola*-infected material.

Unlike PpzA-3, the A_1-domain NRPS code of Esi_PpzA-4 is identical to that of Efe_PpzA-2 (Fig. 2 and *SI Appendix*, Fig. S1) and was therefore predicted to bind L-Pro substrate. However, while *Esi_ppzA-4* is derived from *Efe_ppzA-2*, much of the M_2-domain-encoding region of *Esi_ppzA-4* has been replaced with sequence from an *Efe_ppzA-1* donor (Fig. 2 and *SI Appendix*, Fig. S3). This suggested that Esi_PpzA-4 may be a dedicated **2b** synthetase, as much of the weakly functional Efe_PpzA-2 M_2-domain was replaced. This functionality was confirmed by analysis of *Epichloë*-infected plant material, which showed that high concentrations of **2b** were exclusive to *E. siegelii*-infected samples (Table 1 and *SI Appendix*, Table S3). Furthermore, the synthetic allele *ppzA-S1* can be considered an analog of *Esi_ppzA-4*, and PpzA-S1 was also shown to be a dedicated **2b** synthetase (Fig. 3*D*). However, unlike PpzA-S1, Esi_PpzA-4 retains all of the conserved PpzA-2–conserved substitutions located at the N-terminal end of the M_2 domain (Fig. 2 and *SI Appendix*, Fig. S3). Given that the M_2 domain of Esi_PpzA-4 appears to be fully functional, this implies a causative role for one or both of the PpzA-2–conserved M_2-domain substitutions that are replaced in Esi_PpzA-4 (D1935A and P2003A; *SI Appendix*, Fig. S3).

PPZ-1,4-Dione Synthesis Is Proposed to Occur through Nonenzymatic Dipeptidyl-Thioester Cyclization. In the absence of the R_2 domain, it is conceivable that release of PPZ-1,4-diones from truncated PpzA variants could be catalyzed by separate protein, such as a type II thioesterase. However, candidate genes for this function are not found coclustered with any *ppzA* variant (2, 16, 23). Furthermore, expression of these truncated *ppzA* variants alone was sufficient to achieve PPZ-1,4-dione production in the naive host *P. paxilli* at levels comparable to **1a** production by PN2013/*Efe_ppzA-1* strains (Figs. 3*D* and 4 and *SI Appendix*, Table S1). Given that transcription of all *ppzA* variants in *P. paxilli* was controlled by identical regulatory sequences, this suggests that all PpzA variants have similar biosynthetic efficiencies. This would therefore require a hypothetical endogenous *P. paxilli* protein that can catalyze dipeptidyl release at a rate similar to the integrated R_2 domain of PpzA-1, which seems unlikely. We therefore propose that PPZ-1,4-dione production by PpzA proteins occurs through a spontaneous nucleophilic substitution reaction where

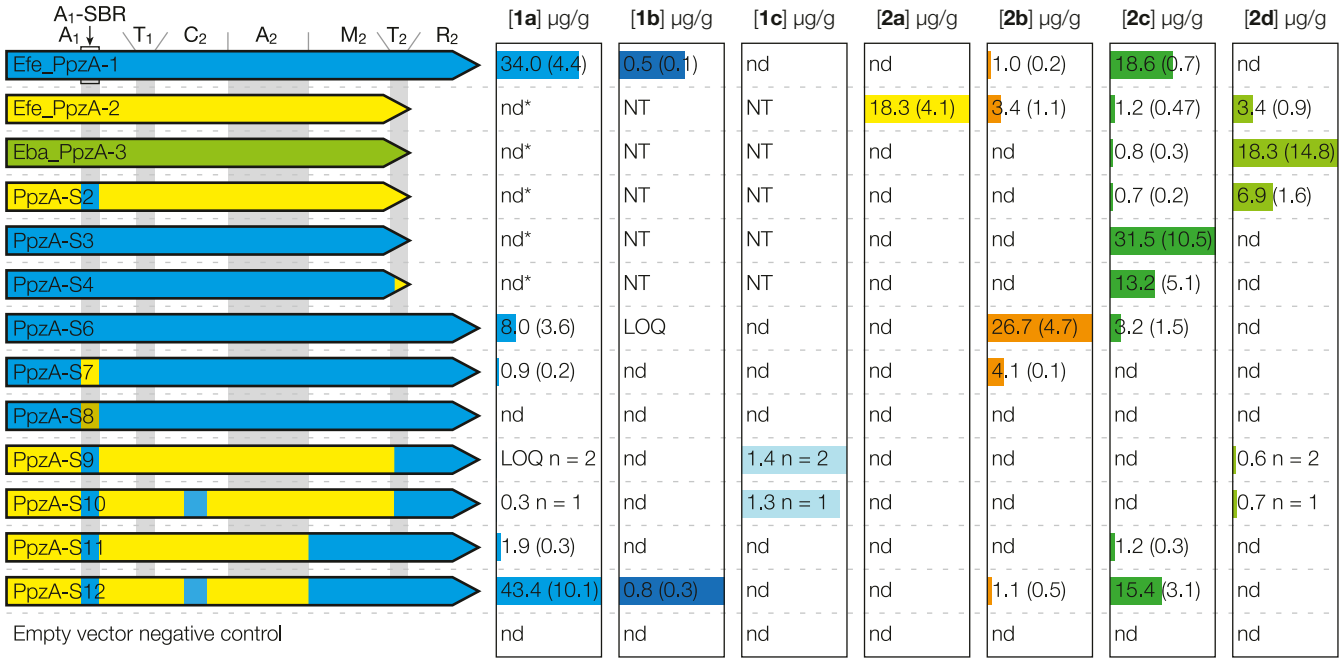

Fig. 4. PPZ synthesis profiles of different PpzA proteins. Each row shows the PPZ profile of T4HP-fed *P. paxilli* strains expressing each of the proteins shown to the left of the image. Protein maps are annotated with domain boundaries, and synthetic hybrid proteins are colored to indicate the natural PpzA protein from which each region is derived. Concentrations are averaged across at least 3 independent transformants for each *ppzA* expression construct, with the SEM shown in parentheses. Where PPZ production was not observed for all strains, the number (n) of strains used to generate the listed concentration is shown instead of the SE. Metabolites that were not detected are annotated "nd" (limit of detection 0.05 μg/g). LOQ indicates detection of a metabolite at a concentration below the limit of quantitation (0.2 μg/g). NT = not tested. The **1a** concentrations for a small subset of these analyses, which are indicated by an asterisk (*), were determined using slightly less sensitive method (limit of detection 0.1 μg/g). Colored bars illustrate the relative between-sample concentration for each metabolite. The concentrations listed for **1b–c** are estimates based on the response factor of synthetic **1a**, while the concentrations listed for **2b–d** are estimates based on the response factor of synthetic **2a**.

the dipeptidyl-thioester sulfur atom is replaced by the pyrrolidine nitrogen atom, releasing a product with a 2,5-diketopiperazine (DKP) core (Fig. 5). This mechanism is equivalent to the pathway proposed by Stachelhaus et al. (24) for biosynthesis of the DKP cyclo(Phe, Pro) by an artificially truncated 2-module variant of the gramicidin synthetase NRPS complex and would not require a termination domain.

Interestingly, biosynthesis of the PPZ-1,4-diones **2b** and **2c** was also observed for PpzA-1 proteins (Fig. 4, Table 2, and *SI Appendix*, Table S1), including the PpzA-1 ortholog from the insect pathogen *Metarhizium rileyi*, which belongs to the same family (Clavicipitaceae) as the *Epichloë* genus. This suggests that nonenzymatic cyclization of the dipeptidyl-thioester intermediate is not a novel innovation of the truncated PpzA proteins. Rather, nonenzymatic cyclization appears to compete with R_2-catalyzed reduction for dipeptidyl thioester release from PpzA-1, and both activities were likely present in PpzA-1 from the last common ancestor of the *Epichloë* and *Metarhizium* genera. The transposable element insertion observed in all *ppzA-2*, *ppzA-3*, and *ppzA-4* alleles could therefore have eliminated the competing reductive release pathway through deletion of the R_2-encoding sequence, resulting in immediate subfunctionalization of the encoded PpzA protein into a dedicated PPZ-1,4-dione synthetase.

To investigate this hypothesis, synthetic allele *ppzA-S3* was generated by truncating *Efe_ppzA-1* to the same length as *Efe_ppzA-2*. This R_2-domain deletion abolished production of **1a** but did not affect production of **2c** (Fig. 4). Exchanging the T_2-encoding sequence of *ppzA-S3* for its *Efe_ppzA-2* equivalent did not further improve **2c** production (PpzA-S4; Fig. 4), nor did exchanging the *Efe_ppzA-2* T_2-encoding sequence for its *Efe_ppzA-1* equivalent reduce production of **2a** (PpzA-S5; Fig. 3*D*). These results demonstrate that **2c** release from PpzA proteins occurs through an R_2-independent mechanism, with R_2-domain truncation resulting in immediate subfunctionalization of PpzA-1 into a dedicated **2c** synthetase. Furthermore, although PpzA-2–conserved substitutions were identified within the T_2 domain (Fig. 1 and *SI Appendix*, Fig. S4), these impart no specific contribution to the efficiency of DKP formation. It was also observed that the artificially truncated *ppzA-S3* allele is similar to 1 of the 2 *ppzA* alleles present in *Epichloë uncinata* strain e167, which is an asexual allopolyploid hybrid species of *E. bromicola* and *E. typhina* subsp. *poae* (22, 25). Both *E. uncinata* alleles were previously thought to be pseudogenes (14); however, the premature stop codon in the *E. typhina*-derived allele only affects translation of the R_2 domain. We therefore predicted this "*Eun_ppzA-5*" allele may encode a **2c** synthetase analogous to PpzA-S3. Analysis of *Epichloë*-infected plant material confirmed that *E. uncinata*-infected material contains **2c**, and this was the only plant material tested that contained **2c** in the absence of **1a** (Table 1). These results support the hypothesis that Eun_PpzA-5 is a dedicated **2c** synthetase, meaning that subfunctionalization of PpzA-1 into a dedicated PPZ-1,4-dione synthetase appears to have occurred at least twice in the evolutionary history of *Epichloë*.

The rate of nonenzymatic dipeptidyl-thioester cyclization would depend on the strength of the attacking nitrogen nucleophile, and the pyrrolidine hydroxy group could reduce this nucleophilicity in 4HP relative to Pro. To investigate if nonenzymatic cyclization of Pro-meArg thioesters is more efficient than for 4HP-meArg thioesters, synthetic allele *ppzA-S6* was generated by PCR-induced mutagenesis of *Efe_ppzA-1* to replace 3 A_1 SBR NRPS code residues of PpzA-1 with their Pro-binding PpzA-2 equivalents (Fig. 1*B* and *SI Appendix*, Fig. S1). This significantly reduced production of **1a** by PpzA-S6 relative to Efe_PpzA-1, and PN2013/*ppzA-S6* cultures instead produced substantial quantities of **2b** (Fig. 4). Levels of **1a** in PN2013/*ppzA-S6* cultures were also still strongly inducible through T4HP feeding (*SI Appendix*,

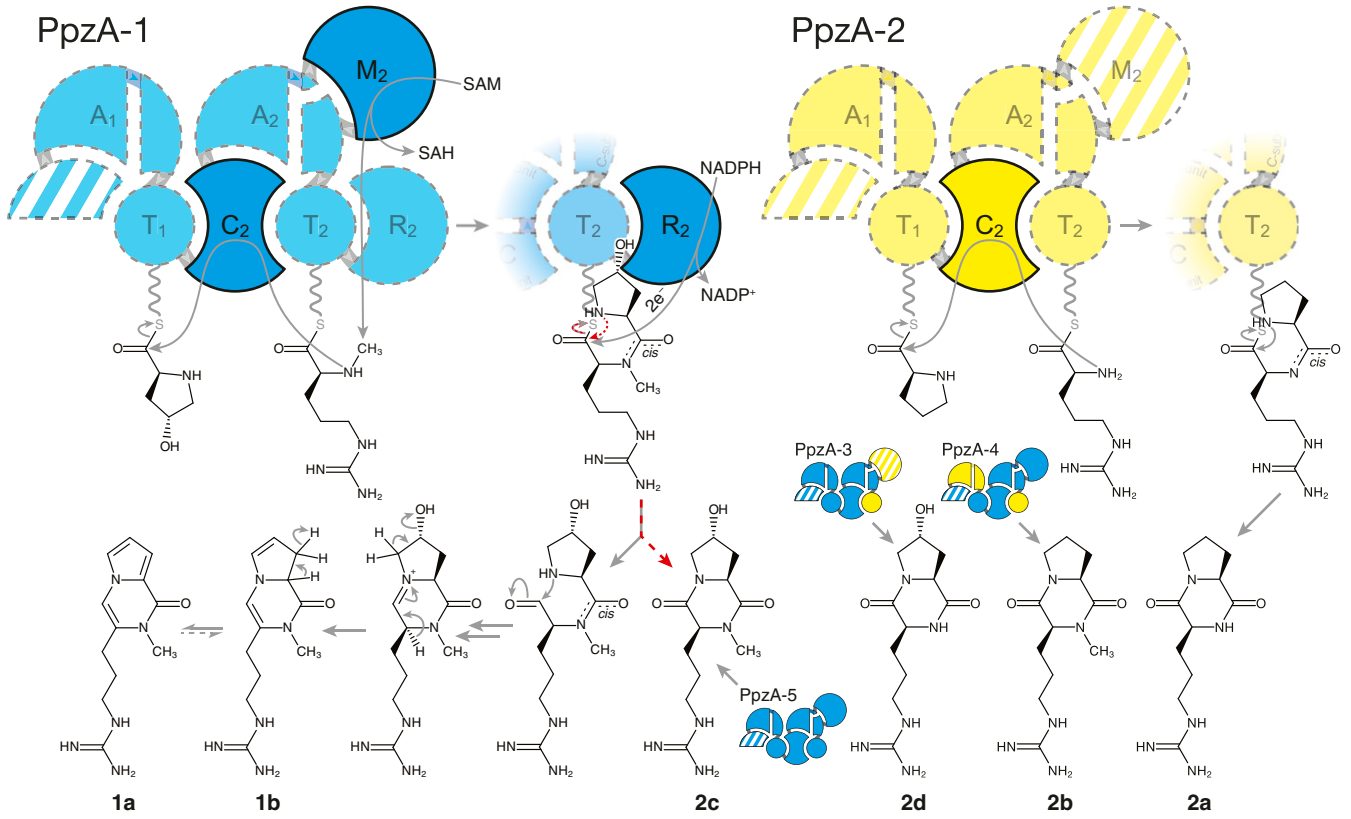

Fig. 5. Proposed PPZ biosynthetic pathways. Predicted reactions are shown after amino acid substrates have been selected, activated, and thio-tethered. Hatched shading is used to indicate the noncatalytic N-terminal partial C domain (13) and the weakly or nonfunctional PpzA-2 M_2-domain. Residual M_2-domain activity means that some PpzA-2 proteins produce both **2a** and **2b**. T4HP is shown as the A_1-domain substrate for PpzA-1 and PpzA-3, but feeding experiments suggest this may be replaceable with C4HP. Biosynthesis of **2d** by PpzA-3 is equivalent to biosynthesis of **2a** by PpzA-2, except the L-Pro substrate is replaced with 4HP. Biosynthesis of **2b** by PpzA-4 is equivalent to PpzA-2-catalyzed biosynthesis of **2a**, except the M_2 domain of PpzA-4 is functional. In the absence of the R_2 domain, PpzA-5 biosynthesis of **2c** occurs via the competing autocatalytic dipeptidyl release pathway shown in red for PpzA-1.

Table S1), indicating that incomplete A_1-domain binding specificity conversion rather than oxidation of a Pro-incorporating intermediate was the source of this residual **1a** biosynthesis. Furthermore, no metabolite corresponding to a hypothetical Pro-derived reduced analog of **1a** ($C_{12}H_{21}N_5O$, 252.18 *m/z*) was detected in any PN2013/*ppzA-S6* samples. These results suggest that nonenzymatic cyclization of Pro-incorporating dipeptidyl thioesters occurs much more rapidly than R_2-catalyzed reductive release from PpzA proteins. In contrast, nonenzymatic release of 4HP-incorporating dipeptidyl thioesters appears to be slower, allowing R_2-catalyzed reductive release to dominate. However, an alternative hypothesis where Pro-incorporating dipeptidyl intermediates are incompetent as R_2-domain substrates cannot be excluded.

Attempts to further reduce 4HP binding specificity were made by exchanging the A_1 SBR-encoding region of *Efe_ppzA-1* for that of *Efe_ppzA-2* or *Ety_ppzA-2* to generate synthetic alleles *ppzA-S7* and *ppzA-S8*, respectively. However, concentrations of both **1a** and **2b** were significantly reduced in PN2013/*ppzA-S7* cultures, while PN2013/*ppzA-S8* cultures did not produce detectable levels of any product (Fig. 4), indicating these subdomain exchanges were detrimental to protein function. This suggests that targeted mutagenesis of NRPS code residues may be a superior strategy to subdomain swaps when attempting to modify binding specificity toward an amino acid that is closely related to the cognate substrate for that A domain.

Restoring 1a Biosynthesis to Efe_PpzA-2 Defines All Functionally Divergent Regions. We have shown that the Efe_PpzA-2 A_1 SBR and M_2 domains are functionally divergent compared to PpzA-1 proteins, and that the R_2 domain is required for **1a** biosynthesis. However, it was not clear if all functionally divergent regions within Efe_PpzA-2 had been identified. To investigate this, synthetic hybrid alleles *ppzA-S9* through *ppzA-S12* were generated by iterative replacement of the divergent *Efe_ppzA-2* A_1 SBR, C_2-, M_2-, and R_2-encoding regions with their *Efe_ppzA-1* equivalents. These alleles were expressed in *P. paxilli* PN2013 and assayed for **1a** production. Replacement of the A_1 SBR and R_2-encoding sequences alone in PpzA-S9 did not result in synthesis of **1a** or **2c**, although small amounts of the nonmethylated analogs **1c** and **2d** were detected in some strains (Fig. 4 and *SI Appendix, Supplementary Note*). Equivalent results were observed when the divergent C_2-encoding region was additionally replaced in PpzA-S10 (Fig. 4). Replacement of the A_1 SBR, M_2-, and R_2-encoding sequences in PpzA-S11 resulted in low but consistent production of **1a** (Fig. 4), and the additional replacement of the divergent portion of the C_2-encoding region in PpzA-S12 fully restored **1a** production to the same level as wild-type Efe_PpzA-1 (Fig. 4). This shows that most Efe_PpzA-2–derived regions in PpzA-S12 (55%) are functionally equivalent to their Efe_PpzA-1 counterparts, although the C_2 domain may have functionally diverged. However, no PpzA-2–conserved substitutions are present within the C_2 domain (Fig. 1*A*), and modeling the Efe_PpzA-1 and Efe_PpzA-2 C_2-domain structures did not reveal any clustering of differential residues (*SI Appendix*, Fig. S8). The cause of this apparent differentiation between Efe_PpzA-1 and Efe_PpzA-2 C_2-domains therefore remains unclear.

Discussion

Fungi typically contain a large number of secondary metabolism gene clusters, each of which encodes proteins involved in biosynthesis of a specific class of natural products (26). Within-class natural product diversity is usually a function of gene gain or loss polymorphisms within these clusters (27), as has been well documented for *Epichloë* species (16, 28). It is understandable that the PPZ product diversity encoded by the orphaned *ppzA* gene in *Epichloë* spp. was thought to be restricted to **1a**, especially given that *PPZ* clusters containing *ppzA-1* orthologs alongside at least 6 accessory genes were recently identified in *Metarhizium* and *Cladonia* spp. (23). However, the results presented here show that allelic variants of a single NRPS can generate metabolite diversity comparable to many gene clusters, with 4 functionally distinct *ppzA* alleles described here in addition to the previously characterized *ppzA-1* (2, 23). Of these, the *ppzA-2*, *ppzA-3*, and *ppzA-4* alleles all derive from the same 3′ sequence deletion caused by a transposable element insertion (14, 15, 18), whereas *ppzA-5* contains a premature stop codon near the 3′ end of the gene (14). In the absence of the C-terminal R_2 domain, these truncated PpzA proteins are proposed to utilize nonenzymatic dipeptidyl-thioester cyclization to release a range of DKP-containing PPZ-1,4-dione products. Functional divergence of some PpzA-2 domains relative to their PpzA-1 progenitors is also demonstrated; PpzA-2 A_1 domains have evolved to bind L-Pro as substrate instead of 4HP, and PpzA-2 proteins produce nonmethylated products due to their weakly active or inactive M_2 domains. Recombination between *ppzA* alleles is also shown to have further enhanced the PPZ diversity encoded by this locus.

Fungi are known to produce a huge variety of DKP metabolites (29), and previous studies have described a number of fungal NRPS pathways dedicated to DKP biosynthesis (1, 30–35). While the mechanism of DKP formation in many of these NRPS pathways has not yet been resolved, this could conceivably occur through nonenzymatic cyclization of the dipeptidyl thioester. However, spontaneous *trans* → *cis* isomerization of the dipeptidyl peptide bond would be required before cyclization could occur, which is usually suppressed by steric repulsions between consecutive C^{α}-linked side chains (36). Gao et al. (37) demonstrated that fungal NRPSs can utilize a C-terminal condensation-like "C_T" domain to catalyze peptidyl cyclization, and all of the dedicated DKP-synthesizing fungal NRPSs described to date terminate with a C-like domain that could conceivably catalyze DKP release (1, 30–35). Indeed, Baccile et al. (38) recently demonstrated that the 2-module NRPS GliP requires a C-terminal C_T–T_3 didomain to catalyze synthesis of the DKP cyclo(Phe, Ser), which is the first step in gliotoxin biosynthesis, and proposed that this is a general mechanism for DKP biosynthesis by fungal NRPSs. Nonenzymatic dipeptidyl cyclization has been observed for some larger NRPS systems when peptidyl chain elongation is stalled by downstream substrate starvation, incorrect substrate loading, or artificial protein truncation (24, 39, 40). Formation of these DKP side-products typically involves Xaa-Pro (Xaa = any amino acid) or *N*-methylated "tertiary" dipeptidyl intermediates, as the influence of a third *N*-linked carbon atom in these tertiary peptide bonds significantly reduces the energy differential between the *cis* and *trans* isomers. However, spontaneous *trans* → *cis* isomerization of tertiary dipeptidyl peptide bonds still appears to be considerably slower than other NRPS-catalyzed reactions (41). Biosynthesis of the PPZ-1,4-diones by PpzA proteins therefore presents an enigma, as **2a** and **2d** biosynthesis does not occur via a tertiary dipeptidyl intermediate. These biosynthetic pathways should therefore be severely bottlenecked by *trans* → *cis* isomerization, yet our results suggest that the efficiency of PpzA-2-catalyzed **2a** biosynthesis is comparable to that of PpzA-1–catalyzed **1a** biosynthesis. This suggests that PpzA proteins are able to predispose the dipeptidyl thioester toward the *cis* conformation, for example by utilizing a C_2 domain that specifically catalyzes *cis* peptide bond formation. While such a mechanism would obviously benefit PPZ-1,4-dione biosynthesis by removing the peptide isomerization bottleneck, it could also ensure that reductively released dipeptide-aldehyde intermediates are rapidly cyclized during **1a** biosynthesis, inhibiting potentially harmful interactions with external nucleophiles. Nonenzymatic dipeptidyl cyclization is likely also facilitated by the extraordinarily high nucleophilicity of the Pro-derived nitrogen atom in Pro–Arg dipeptidyls (42), which may be reduced by the 4-hydroxy group of 4HP. This could explain why 4HP–meArg dipeptidyl intermediates are primarily—but not exclusively—released via R_2-catalyzed reduction during **1a** biosynthesis, whereas nonenzymatic cyclization completely dominated release of Pro-containing dipeptidyl intermediates even when a functional R_2 domain was present.

Our results show that very few mutations would have been required during the transition of PpzA-1 to PpzA-2. Interestingly, the A_1-domain Pro binding, nonmethylated product synthesis, and nonenzymatic dipeptidyl cyclization activities that characterize PpzA-2 proteins were also observed to be reactions that are weakly catalyzed by PpzA-1 proteins. PpzA-2 proteins therefore appear to have evolved through a series of domain subfunctionalizations that cumulatively resulted in an NRPS that efficiently synthesizes **2a**, with our results suggesting this evolutionary cascade could have been triggered by the transposon-mediated R_2-encoding sequence deletion observed in all *ppzA-2* alleles (14, 15, 18). We also showed that further PPZ diversity was generated in the *E. festucae* and *E. bromicola* lineages by recombination-mediated shuffling of functionally divergent sequences between *ppzA-1* and *ppzA-2* alleles. This generated the *ppzA-3* and *ppzA-4* alleles, each of which encodes a protein with novel biosynthetic activity. Due to their modular nature, recombination-vectored domain exchange has long been proposed as a dominant mechanism driving NRPS evolution and diversity, although relatively few examples have been described (43–45). The *ppzA* loci of *Epichloë* spp. represent a remarkably clean example of this process, as even the most divergent alleles share ≥95% DNA sequence identity. Interestingly, *Ebr_ppzA-2* and *Efe_ppzA-2* were also shown to have undergone extensive recombination, yet these alleles remain functionally equivalent to their nonrecombinant *Ety_ppzA-2* counterparts. This suggests that the *E. bromicola* and *E. festucae* lineages may have inherited a nonfunctional *ppzA-2* pseudogene that was repaired through recombination with a functional *ppzA-1* donor. Surprisingly, the equivalent recombination patterns exhibited by *Ebr_ppzA-2*/*Efe_ppzA-2* and *Ebr_ppzA-3*/*Efe_ppzA-3* alleles appear to have arisen after divergence of the *E. bromicola* and *E. festucae* lineages in an example of convergent evolution.

Interestingly, the *ppzA-2*–derived regions of the recombinant *ppzA-2*, *ppzA-3*, and *ppzA-4* alleles exhibit TSP, meaning that the donor *ppzA-2* alleles diverged from *ppzA-1* before emergence of the *E. bromicola*, *E. festucae*, and *E. typhina* lineages. Each of these species also has conspecific strains with functionally distinct *ppzA* alleles, suggesting that the *ppzA* locus is subject to negative frequency-dependent or "balancing" selective pressures that reduce the fitness of the most common allele. The major histocompatibility complex (MHC) in mammals is a well-characterized example of TSP, with balancing selection driven by pathogenic coevolution thought to have prevented fixation of any one MHC allele (46, 47). Analogous balancing selective pressures on the *ppzA* locus could be driven by coevolution of resistance against the dominant PPZ chemotype in target insects, or by displacement of susceptible insect species by resistant competitors. For example, previous experiments have demonstrated that although **1a** exhibits potent feeding deterrent activity against *L. bonariensis* it is not active against several other species (12, 48, 49), which could potentially be susceptible to other PPZs. The *N*-methyl group is known to be essential for the bioactivity of **1a** (12), suggesting that while *N*-methylated PPZ-1,4-diones **2b** and **2c** may exhibit

activities similar to **1a**, the nonmethylated PPZ-1,4-diones **2a** and **2d** likely do not. This is supported by the results of Rowan et al. (12), who coincidentally tested **2a** due to its structural analogy with **1a** and demonstrated that **2a** does not deter feeding by *L. bonariensis*. However, **2a** has previously been shown to be a potent inhibitor of chitinase activity (50, 51) and might thus interfere with insect ecdysis or suppress fungal competitors instead. Future studies into **2a–d** would be useful to determine if these PPZ-1,4-diones exhibit agriculturally relevant bioactivities.

The results presented here demonstrate that nonenzymatic dipeptidyl thioester cyclization can be utilized by NRPSs in vivo for the efficient synthesis of DKP-containing metabolites, meaning that this mechanism should be considered when encountering NRPSs without an obvious termination domain. Further studies of PpzA proteins could reveal if these enzymes exhibit features that facilitate this process, such as the proposed *cis*-peptide-forming C domain. Although reports of allelic neofunctionalization and TSP in fungi are rare, we demonstrate that these processes can generate considerable diversity and suggest that both processes may be more prevalent than currently realized. These processes may be particularly relevant at allelic loci where gene conversion is inhibited by the insertion of large homology-breaking DNA elements such as transposons. TSP and lateral gene transfer can also produce similar phylogenetic patterns, particularly between closely related species, meaning that both hypotheses need to be considered carefully.

Materials and Methods

A comprehensive description of the materials and methods used in this study is available in *SI Appendix*. Homologous recombination was used to generate *E. festucae* Δ*ppzA-2* strains. Constructs for *ppzA* expression in *P. paxilli* were assembled with all *ppzA* variants placed under transcriptional control of the same regulatory sequences from the native *P. paxilli* secondary metabolism gene *paxM*. These constructs were introduced into the *P. paxilli* genome via nontargeted integration, with RT-PCR used to select at least 3 independent transformants expressing each *ppzA* variant. Cultures for metabolite analysis were grown for 6 d under standardized conditions, with substrate feeding performed at 4 and 5 d postinoculation. PPZ metabolites were extracted from lyophilized *P. paxilli* mycelia or *Epichloë*-infected plant material and were analyzed using hydrophilic-interaction chromatography-coupled positive electrospray ionization mass spectroscopy (LCMS).

Data Availability. DNA sequence data generated during this study have been deposited in the GenBank database under accession numbers MN605951 to MN605962. Raw PPZ concentrations from all LCMS analyses, MS^n spectra for all PPZ metabolites, and NMR spectra for **2a** are available in *SI Appendix*.

ACKNOWLEDGMENTS. Work in the H.B.B. laboratory was funded in part by the LOEWE Centre for Translational Biodiversity Genomics. D.B. was supported by a Massey University PhD scholarship from Massey University and funding from the New Zealand Tertiary Education Commission provided through the Bioprotection Research Center. B.S. was supported by an Alexander von Humboldt Research Award. We thank Shaun Bushman (US Department of Agriculture) and Devish Singh (Barenbrug) for providing access to field trials for sampling and Dr. Patrick Edwards (Massey University) for assistance in generating NMR data.

1. B. Gu, S. He, X. Yan, L. Zhang, Tentative biosynthetic pathways of some microbial diketopiperazines. *Appl. Microbiol. Biotechnol.* **97**, 8439–8453 (2013).
2. A. Tanaka, B. A. Tapper, A. Popay, E. J. Parker, B. Scott, A symbiosis expressed non-ribosomal peptide synthetase from a mutualistic fungal endophyte of perennial ryegrass confers protection to the symbiotum from insect herbivory. *Mol. Microbiol.* **57**, 1036–1050 (2005).
3. M. A. Wyatt *et al*., *Staphylococcus aureus* nonribosomal peptide secondary metabolites regulate virulence. *Science* **329**, 294–296 (2010).
4. M. Zimmermann, M. A. Fischbach, A family of pyrazinone natural products from a conserved nonribosomal peptide synthetase in *Staphylococcus aureus*. *Chem. Biol.* **17**, 925–930 (2010).
5. R. D. Süssmuth, A. Mainz, Nonribosomal peptide synthesis-principles and prospects. *Angew. Chem. Int. Ed. Engl.* **56**, 3770–3821 (2017).
6. C. T. Walsh, Insights into the chemical logic and enzymatic machinery of NRPS assembly lines. *Nat. Prod. Rep.* **33**, 127–135 (2016).
7. C. T. Walsh, R. V. O'Brien, C. Khosla, Nonproteinogenic amino acid building blocks for nonribosomal peptide and hybrid polyketide scaffolds. *Angew. Chem. Int. Ed. Engl.* **52**, 7098–7124 (2013).
8. L. Du, L. Lou, PKS and NRPS release mechanisms. *Nat. Prod. Rep.* **27**, 255–278 (2010).
9. D. D. Rowan, M. B. Hunt, D. L. Gaynor, Peramine, a novel insect feeding deterrent from ryegrass infected with the endophyte *Acremonium loliae*. *J. Chem. Soc. Lond. Chem. Commun.* **12**, 935–936 (1986).
10. C. L. Schardl, The epichloae, symbionts of the grass subfamily *Pooideae*. *Ann. Mo. Bot. Gard.* **97**, 646–665 (2010).
11. R. Prestidge, G. Barker, R. Pottinger, "The economic cost of Argentine stem weevil in pastures in New Zealand" in *Proceedings of the 44th New Zealand Weed and Pest Control Conference* (New Zealand Weed and Pest Control Society Inc., Auckland, New Zealand, 1991), pp. 165–170.
12. D. D. Rowan, J. J. Dymock, M. A. Brimble, Effect of fungal metabolite peramine and analogs on feeding and development of argentine stem weevil (*Listronotus bonariensis*). *J. Chem. Ecol.* **16**, 1683–1695 (1990).
13. D. Kalb, G. Lackner, M. Rappe, D. Hoffmeister, Activity of α-aminoadipate reductase depends on the N-terminally extending domain. *ChemBioChem* **16**, 1426–1430 (2015).
14. D. Berry *et al*., Disparate independent genetic events disrupt the secondary metabolism gene *perA* in certain symbiotic *Epichloë* species. *Appl. Environ. Microbiol.* **81**, 2797–2807 (2015).
15. D. J. Fleetwood *et al*., Abundant degenerate miniature inverted-repeat transposable elements in genomes of epichloid fungal endophytes of grasses. *Genome Biol. Evol.* **3**, 1253–1264 (2011).
16. C. L. Schardl *et al*., Plant-symbiotic fungi as chemical engineers: Multi-genome analysis of the clavicipitaceae reveals dynamics of alkaloid loci. *PLoS Genet.* **9**, e1003323 (2013).
17. I. K. Hettiarachchige *et al*., Genetic modification of asexual *Epichloë* endophytes with the *perA* gene for peramine biosynthesis. *Mol. Genet. Genomics* **294**, 315–328 (2018).
18. C. L. Schardl *et al*., Currencies of mutualisms: Sources of alkaloid genes in vertically transmitted epichloae. *Toxins (Basel)* **5**, 1064–1088 (2013).
19. D. J. Winter *et al*., Repeat elements organise 3D genome structure and mediate transcription in the filamentous fungus Epichloë festucae. *PLoS Genet.* **14**, e1007467 (2018).
20. M. A. Marahiel, T. Stachelhaus, H. D. Mootz, Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chem. Rev.* **97**, 2651–2674 (1997).
21. T. Stachelhaus, H. D. Mootz, M. A. Marahiel, The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6**, 493–505 (1999).
22. K. D. Craven, J. D. Blankenship, A. Leuchtmann, K. Hignight, C. L. Schardl, Hybrid fungal endophytes symbiotic with the grass *Lolium pratense*. *Sydowia* **53**, 44–73 (2001).
23. D. Berry *et al*., Orthologous peramine and pyrrolopyrazine-producing biosynthetic gene clusters in *Metarhizium rileyi*, *Metarhizium majus* and *Cladonia grayi*. *Environ. Microbiol.* **21**, 928–939 (2019).
24. T. Stachelhaus, H. D. Mootz, V. Bergendahl, M. A. Marahiel, Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain. *J. Biol. Chem.* **273**, 22773–22781 (1998).
25. A. Leuchtmann, C. W. Bacon, C. L. Schardl, J. F. White, Jr, M. Tadych, Nomenclatural realignment of Neotyphodium species with genus Epicholë. *Mycologia* **106**, 202–215 (2014).
26. A. Rokas, J. H. Wisecaver, A. L. Lind, The birth, evolution and death of metabolic gene clusters in fungi. *Nat. Rev. Microbiol.* **16**, 731–744 (2018).
27. A. L. Lind *et al*., Drivers of genetic diversity in secondary metabolic gene clusters within a fungal species. *PLoS Biol.* **15**, e2003583 (2017).
28. C. A. Young *et al*., Genetics, genomics and evolution of ergot alkaloid diversity. *Toxins (Basel)* **7**, 1273–1302 (2015).
29. X. Wang, Y. Li, X. Zhang, D. Lai, L. Zhou, Structural diversity and biological activities of the cyclodipeptides from fungi. *Molecules* **22**, 2026 (2017).
30. W.-B. Yin, A. Grundmann, J. Cheng, S.-M. Li, Acetylaszonalenin biosynthesis in *Neosartorya fischeri*. Identification of the biosynthetic gene cluster by genomic mining and functional proof of the genes by biochemical investigation. *J. Biol. Chem.* **284**, 100–109 (2009).
31. K. Mundt, B. Wollinsky, H. L. Ruan, T. Zhu, S. M. Li, Identification of the verruculogen prenyltransferase FtmPT3 by a combination of chemical, bioinformatic and biochemical approaches. *ChemBioChem* **13**, 2583–2592 (2012).
32. S. Maiya, A. Grundmann, S. M. Li, G. Turner, The fumitremorgin gene cluster of Aspergillus fumigatus: Identification of a gene encoding brevianamide F synthetase. *ChemBioChem* **7**, 1062–1069 (2006).
33. C. García-Estrada *et al*., A single cluster of coregulated genes encodes the biosynthesis of the mycotoxins roquefortine C and meleagrin in *Penicillium chrysogenum*. *Chem. Biol.* **18**, 1499–1512 (2011).
34. T. Correia, N. Grammel, I. Ortel, U. Keller, P. Tudzynski, Molecular cloning and analysis of the ergopeptine assembly system in the ergot fungus *Claviceps purpurea*. *Chem. Biol.* **10**, 1281–1292 (2003).
35. C. J. Balibar, C. T. Walsh, GliP, a multimodular nonribosomal peptide synthetase in *Aspergillus fumigatus*, makes the diketopiperazine scaffold of gliotoxin. *Biochemistry* **45**, 15029–15038 (2006).
36. D. E. Stewart, A. Sarkar, J. E. Wampler, Occurrence and role of *cis* peptide bonds in protein structures. *J. Mol. Biol.* **214**, 253–260 (1990).
37. X. Gao *et al*., Cyclization of fungal nonribosomal peptides by a terminal condensation-like domain. *Nat. Chem. Biol.* **8**, 823–830 (2012).
38. J. A. Baccile *et al*., Diketopiperazine formation in fungi requires dedicated cyclization and thiolation domains. *Angew. Chem. Int. Ed. Engl.* **58**, 14589–14593 (2019).

39. A. W. Schultz *et al*., Biosynthesis and structures of cyclomarins and cyclomarazines, prenylated cyclic peptides of marine actinobacterial origin. *J. Am. Chem. Soc.* **130**, 4507–4516 (2008).
40. S. G. Lee, F. Lipmann, Isolation of amino acid activating subunit–pantetheine protein complexes: Their role in chain elongation in tyrocidine synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 2343–2347 (1977).
41. S. Gruenewald, H. D. Mootz, P. Stehmeier, T. Stachelhaus, In vivo production of artificial nonribosomal peptide products in the heterologous host *Escherichia coli*. *Appl. Environ. Microbiol.* **70**, 3282–3291 (2004).
42. F. Brotzel, H. Mayr, Nucleophilicities of amino acids and peptides. *Org. Biomol. Chem.* **5**, 3814–3820 (2007).
43. M. Crüsemann, C. Kohlhaas, J. Piel, Evolution-guided engineering of nonribosomal peptide synthetase adenylation domains. *Chem. Sci. (Camb.)* **4**, 1041–1045 (2013).
44. D. P. Fewer *et al*., Recurrent adenylation domain replacement in the microcystin synthetase gene cluster. *BMC Evol. Biol.* **7**, 183 (2007).
45. B. Mikalsen *et al*., Natural variation in the microcystin synthetase operon *mcyABC* and impact on microcystin production in *Microcystis* strains. *J. Bacteriol.* **185**, 2774–2785 (2003).
46. J. Klein, Origin of major histocompatibility complex polymorphism: The trans-species hypothesis. *Hum. Immunol.* **19**, 155–162 (1987).
47. S. Sommer, The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Front. Zool.* **2**, 16 (2005).
48. O. J. Ball *et al*., Importance of host plant species, *Neotyphodium* endophyte isolate, and alkaloids on feeding by *Spodoptera frugiperda* (Lepidoptera: Noctuidae) larvae. *J. Econ. Entomol.* **99**, 1462–1473 (2006).
49. O. J. P. Ball, C. O. Miles, R. A. Prestidge, Ergopeptine alkaloids and *Neotyphodium lolii*-mediated resistance in perennial ryegrass against adult *Heteronychus arator* (Coleoptera: Scarabaeidae). *J. Econ. Entomol.* **90**, 1382–1391 (1997).
50. H. Izumida, N. Imamura, H. Sano, A novel chitinase inhibitor from a marine bacterium, *Pseudomonas* sp. *J. Antibiot. (Tokyo)* **49**, 76–80 (1996).
51. H. Izumida, M. Nishijima, T. Takadera, A. M. Nomoto, H. Sano, The effect of chitinase inhibitors, cyclo(Arg-Pro) against cell separation of *Saccharomyces cerevisiae* and the morphological change of *Candida albicans*. *J. Antibiot. (Tokyo)* **49**, 829–831 (1996).
52. V. N. Minin, K. S. Dorman, F. Fang, M. A. Suchard, Dual multiple change-point model leads to more accurate recombination detection. *Bioinformatics* **21**, 3034–3042 (2005).