

Reconstructing the evolution of a key fungal isocyanide megasynthase using genomes of lichenized fungi

Grant R. Nickles¹, Anna Vaiana², Sung Chul Park¹, Karen Broz³, Harrison P. Estes¹, Theo Llewellyn⁴, Milton T. Drott^{3,5}, Nancy P. Keller^{1,6,#}, Garima Singh^{2,7,#}

1 Department of Medical Microbiology and Immunology, University of Wisconsin—Madison, Madison, Wisconsin, USA

2 Department of Biology, University of Padova, Via U. Bassi, 58/B, 35121 Padova, Italy

3 Cereal Disease Laboratory, Agricultural Research Service, US Department of Agriculture, St. Paul, Minnesota, USA

4 Department of Life Sciences, Imperial College London, South Kensington Campus London, UK

5 Department of Plant Pathology, University of Minnesota, St. Paul, Minnesota, USA

6 Department of Plant Pathology, University of Wisconsin—Madison, Madison, Wisconsin, USA

7 National Biodiversity Future Center (NBFC), Piazza Marina, 61, 90133, Palermo, Italy

corresponding authors (Nancy Keller: npkeller@wisc.edu; Garima Singh: garima.singh@unipd.it)

Summary

The products of isocyanide synthase (ICS) biosynthetic gene clusters (BGCs) have been implicated in microbial interactions, pathogenesis, and metal homeostasis. While several ICS BGCs have been described as mediating metal-associated ecologies, the evolutionary history of this recently described class of secondary metabolite is unexplored among Lecanoromycetes, a clade comprised predominantly of symbiotic lichen-forming fungi (LFF), known to thrive in metal-contaminated environments (e.g., mine spills). Analyzing nearly 4,000 fungal genomes, including many Lecanoromycetes, we identified six unique LFF-associated ICS clades. Reconstructing the evolutionary history of LFF ICSs uncovered a widespread “split” variant of the copper-responsive (*crmA*) pathway, where the ICS and NRPS-like components are encoded on separate genes, contrasting the canonical “fused” *crmA* megasynthase. Metabolic characterization and genetic deletions in *Fusarium graminearum* confirmed that this split architecture is functionally equivalent

to the fused form. Phylogenetic reconstruction indicates the fused *crmA* arose from a split ancestor, whose NRPS-like component evolved from a canonical thioester reductase architecture likely via domain replacement. Redefinition of the *crmA* pathway to include the “split” ICS/NRPS-like variant reveals that *crmA* is the most prevalent ICS in the fungal kingdom. Our work demonstrates how exploring understudied fungal lineages can define new specialized metabolism lineages and reshape our understanding of biosynthetic pathways across the fungal kingdom.

Introduction

Lichen-forming fungi (LFF) exist as an obligate symbiotic association with one or more photosynthetic partners and other associated microbiota, forming composite organisms called lichens^{1,2}. Lichens are ubiquitous and are found in diverse environments and on various substrates, such as rocks, soil, and tree bark^{3–5}. The lichenized lifestyle evolved multiple times during the evolution of fungi; therefore, LFFs are present in various taxonomic classes across the fungal tree of life^{6–10}. Lecanoromycetes, are almost exclusively lichenized while a few LFF are also found across Eurotiomycetes, Dothideomycetes, and Lichinomycetes^{8–10}.

Some lichens have the remarkable ability to colonize highly metal-contaminated environments and accumulate elevated metal concentrations in their thalli without apparent toxic effects^{11,12}. Despite this striking resilience of lichens in metal-contaminated environments, the molecular basis of their metal tolerance remains entirely uncharacterized. Till date, no pathways, or specialized metabolites (SMs) have been linked to this capability, representing a fundamental gap in our understanding of how lichens adapt to metal-toxicity-related abiotic stress.

LFF are prolific producers of SMs that mediate important aspects of fungal biology, including ecological and species interactions^{13–15}. Studies show that the lichen biosynthetic landscape differs from that of non-lichenized fungi. For instance, in contrast to non-ribosomal peptide synthetase (NRPSs) being the predominant BGC class in non-lichenized fungi, polyketide synthase (PKSs) are the most common BGCs in LFF^{16–18}. Notably, the lichen PKS produced depside and tridepsides make up 5% to 20% of the dry weight of the thallus and are recent promising source of new antioxidant and antimicrobial drug candidates¹⁹.

Recently, isocyanide synthases (ICSs) were reported as the fifth most common BGC class in fungi¹⁸. Isocyanides produced by these BGCs are highly bioactive compounds that serve important roles in many metal-associated ecological interactions^{20–22}. The ICS BGC, *crmA* (copper responsive metabolite) pathway is a notable example of a fungal isocyanide SM pathway. Products of the *CrmA* enzyme have been characterized in several fungi including *A. fumigatus*, *Alternaria brassicicola*, *Cochliobolus heterostrophus*, *Penicillium commune*, and *Penicillium pinophilum*.^{21,23–25} *CrmA* was first characterized in *A. fumigatus* by our research group²³, where it was shown that the ICS-domain is attached to an adenylation domain, a peptidyl carrier protein domain (T domain),

and a non-canonical transferase domain. We classified as a novel ICS-nonribosomal peptide synthase (NRPS) hybrid enzyme due to the presence of the canonical NRPS A and T domains²³. Notably, the NRPS-like portion lacks a condensation (C) domain, which is required for peptide bond formation during canonical NRPS synthesis. Subsequent work by our lab demonstrated that CrmA generates a valine-derived isocyanide that can either (i) form trans-BGC products with ergot alkaloid precursors resulting in fumivalines, or (ii) contribute to the biosynthesis of fumicicolins, a family of acylated sugar alcohols^{21,24}. These CrmA-associated metabolites were shown to have broad spectrum antibacterial and antifungal activity whose production was mediated by copper starvation^{21,24,25}.

While our group in Nickles et al. (2023) described the presence of fungal ICS within the Eurotiomycetes, Dothideomycetes, Sordariomycetes and Leotiomycetes¹⁸, the original dataset contained only five Lecanoromycetes genomes, the predominant taxonomic class of LFF fungi. Given the common association between fungal and bacterial isocyanides and metal ecology^{21-23,26}, and the lack of reports on LFF Lecanoromycetes ICS evolution and diversity, we aimed to determine whether lichens possess any novel ICS classes that are specific to lichenized fungi and potentially associated with a symbiotic lifestyle. In this study, we trace the evolution and diversity of 1,847 ICS enzymes, with a particular focus on the LFF Lecanoromycetes ICS. We sought to answer the following questions: 1) What is the frequency and diversity of ICS BGCs in Lecanoromycetes? 2) Are there LFF-specific ICSs? And 3) what insights into evolutionary forces driving ICS evolution can be gained by exploring these understudied fungal lineages?

Methods

Database generation and ICS Gene Extraction

Publicly available assembled and annotated fungal genomes ($n = 3,992$) were downloaded from the NCBI database on 12/1/2023 using the NCBI's Dataset tool, version 11.32.1. Taxonomy annotations for every accession were obtained using the Bio.Entrez package within the BioPython v1.79 suite²⁷. Accessions missing taxonomy information are a result of NCBI not storing the correct taxonomy annotations when the database was curated. Protein domain predictions were obtained for every protein using HMMER v3.4 (e-value $\leq 1e-5$)²⁸ with the Pfam database v34²⁹. The annotated genomes were scanned for ICSs based on the presence of the ICS-specific domain PF05141.1 (named 'DIT1_PvcA' in the Pfam database in reference to the *dit* cluster). To filter out the impact of poorly sequenced or assembled genomes, 290 universal fungal Benchmarking Single Copy Orthologs (hereby referred to as BUSCOs)³⁰ were predicted in the genomes (database: fungi_odb10 v5). Species where more than one genome encoded at least one ICS were filtered to only include the genome with the highest BUSCO count (run on protein mode). This filtering reduced the impact of over-sequenced species on the reconstructed ICS phylogeny and helped filter out poorly sequenced or annotated genomes. A full table with the ICS counts, BUSCO counts,

taxonomic information, and NCBI genome accessions can be found in Table S1. The filtered table which only includes those with ICSs used in subsequent analysis can be found in Table S2.

Phylogenetic inference of ICS Phylogeny and tree clade annotation

First, the set of ICS proteins found in the filtered genomes was aligned using MAFFT v7.520 with the '-auto' parameter³¹ and trimmed with a gap threshold of 0.1 using trimAl v1.5³². We constructed the phylogeny using IQTree2 v2.2.0³³ and ran 1,000 ultrafast bootstrap replicates³⁴. The optimal model of sequence evolution was automatically selected using ModelFinder³⁵. All tree visualizations were constructed using either iTOL v7.1³⁶ or ggtree v3.11.0³⁷ and ggtreeExtra v1.12.0³⁷. The ten ICS GCF classifications from Nickles et al. (2023), which are *dit* (group 1, 2, and 3), *xan*, *crm*, and GCF 1, 4, 6, 7, and 8, were overlaid onto the constructed phylogeny by locating the most recent common ancestor (MRCA) of proteins previously classified within the same GCF. GCF 8 was found to fall within a large clade of proteins surrounding the known *crm* clade, and thus the label was removed (lime green clade in Fig. 1). Two heavily lichen associated clades of ICS enzymes were named lichen group 1 (LG1) and lichen group 2 (LG2). See Table S3 for a full list of proteins labeled with GCF classifications from Nickles et al. (2023)¹⁸.

Gene cluster family characterization of LG1 and LG2 ICS enzymes

Tip names that fell within lichen group 1 (LG1) and lichen group 2 (LG2) was pulled. A python script that uses the Biopython²⁷ package was created to extract the five genes on both sides of LG1 and LG2 ICS genes and save the output in GenBank (gbk) format. We leveraged the zol/fei suite³⁸ for comparative genomic and functional annotation analysis of the extracted loci. Zol (default settings) was used to align the extracted LG1 and LG2 loci and infer orthogroups from the homologous loci surrounding LG1 and LG2. The resulting linkages were visualized using the cgc and cgcg tools in the zol/fei suite. Protein domains were labeled as "biosynthetic", "transporter", or "regulator" with the help of the Pfam predictions for each protein and Pfam functional label tables used by the antiSMASH program³⁹. The core surrounding LG1 and LG2 was defined as the most common orthogroups detected in the species with LG1 and LG2 (cutoff >30%).

Reconstruction of the fungal kingdom's species tree

We reconstructed a kingdom-level species tree by running a coalescent model, ASTRAL v5.7.8^{40,41}, on the 290 BUSCO gene trees. All gene trees were constructed using the same maximum likelihood pipeline that was used to create the ICS phylogeny. A species map file was passed into ASTRAL to force multiple isolates within the same species to be monophyletic. The tree was rooted at *Rozella allomyces* (GCA_000442015.1) which is the only publicly available genome from the earliest diverging lineage of the Fungal Kingdom, Cryptomycota⁴².

Phylogenetic inference of CrmA type ICS and NRPS-like domains

The clade containing known CrmA proteins and related monodomain ICS proteins (corresponding to the lime green and dark green clades in Fig. 1) was extracted from the full ICS phylogeny. A single ICS-TauD protein from GCF 6 (*Fusarium equiseti*, CAG7566245.1) was selected as the outgroup to root the tree, as GCF 6 was the closest clade to the extracted CrmA-type group in the unrooted phylogeny (Fig. 1). To locate the protein domain, start and stop positions, hmmsearch was run in '--domtblout' mode against the Pfam database^{28,29} using all extracted CrmA-type protein sequences as queries. The ICS domain sequence was extracted from each protein based on the coordinates reported by hmmsearch for the best-score ICS domain hit (PF05141.1). For constructing the NRPS-like domain phylogeny, sequences from proteins encoded by genes separate from the *crmA* gene were used directly. If the NRPS-like domain originated from a fused CrmA protein, the N-terminal ICS domain sequence was removed, retaining only the C-terminal NRPS-like portion for the alignment.

NRPS-like domain annotation and classification

The identified NRPS-like proteins (from both fused CrmA and split *crmA* loci) were classified into architectural "Types" (e.g., Type A, Type B, Type C) based on their predicted domain content. For proteins originating from fused *crmA* genes, the full-length CrmA protein sequence was used as input for domain prediction. For proteins originating from split *crmA* loci, only the sequence of the separate NRPS-like protein was used. Domain prediction, classification, and visualization were performed using synthaser v1.1.22⁴³, a tool that analyzes megasynthase domain architecture via NCBI Conserved Domain Database searches (CD-Search). Because the signature *crmA*-associated ICS domain (PF05141) and the non-canonical C-terminal transferase domain found in Type A architectures were not included in the default synthaser ruleset for domain labeling, we created a custom ruleset with the web tool provided by the synthaser developers that matched the domain accessions "COG3207" and "pfam05141" to the ICS domain, and "PLN02481" to the transferase.

Strains, media, growth conditions

All the *Fusarium graminearum* PH-1 strains used in this study are listed in Table S4. The strains were grown on potato dextrose broth with 1.6% agar plate (PDBA) and 150 ug/mL of geneticin was used for the neomycin resistant strains (KLB34, 36, 44, 45). For the experiments, conidia were harvested from PDA cultures grown at 25 °C for 7 days by scraping with an L-shaped cell spreader in 10 mL of 0.01% Tween 80 water (MP Biomedicals, LLC Catalog No. 103170). The conidial suspension was then passed through sterile miracloth (Millipore, catalog number. 475855-1R) to remove hyphal fragments.

Specialized metabolite extraction from *F. graminearum*

Wild-type *F. graminearum* (PH-1) and Δ ICS and Δ NRPS mutants were used to analyze the production of the split type *crmA*-derived specialized metabolites. PDBA plates were point-

inoculated with 2.5 million spores in triplicate and incubated at 25 °C for 7 days in the dark. Each plate was lyophilized and extracted with 30 mL of methanol. Crude extracts were weighed after drying the solvents using the rotary evaporator (Buchi rotavapor R-300) and resuspended in 100% methanol at 1 mg/mL.

UHPLC–HRMS/MS analysis

Ultra-high pressure liquid chromatography–high resolution mass spectrometry (UHPLC–HRMS) data were acquired using a Thermo Scientific Q Exactive Orbitrap mass spectrometer coupled to a Vanquish UHPLC operated in both positive and negative ionization modes. All solvents used were of spectroscopic grade. Extracts from the conditions described below were diluted into 1 mg/mL and used as samples for UHPLC– HRMS/MS. A Waters XBridge BEH-C18 column (2.1 × 100 mm, 1.7 µm) was used with acetonitrile (0.05% formic acid) and water (0.05% formic acid) as solvents at a flow rate of 0.2 mL/min. The screening gradient method for the samples is as follows: Starting at 10% organic for 5 min, followed by a linear increase to 90% organic over 20 min, another linear increase to 98% organic for 2 min, holding at 98% organic for 5 min, decreasing back to 10% organic for 3 min, and holding at 10% organic for the final 2 min, for a total of 37 min. A quantity of 10 µL of each sample was injected into the system for the analysis. Each extract was diluted in 1 mg/mL in methanol and 10 ppm of N-formylvaline²¹ was used as a standard.

Structural alignment of the CrmA domains

Protein sequences and PDB files were collected from UniProt⁴⁴ and the AlphaFold Structure Database^{45,46} (accessed May 5, 2025), respectively, for CrmA of *Aspergillus fumigatus*, A0A6G1KVI0 (ICS) and A0A6G1KV88 (NRPS-like region) of *Teratosphaeria nubilosa*, and E4V2M2 (ICS) and E4V2M5 (NRPS-like thioester reductase) of *Arthroderma gypseum*. Visualizations of individual proteins were made using the AlphaFold Protein Structure Database⁴⁷. The sequences were submitted to the InterProScan (v5)⁴⁸ web server to determine coordinates for structural alignment (Table S5). The RCSB PDB Pairwise Alignment tool server⁴⁹ was used with a jFATCAT (flexible) alignment algorithm and default parameters to generate structural alignments and statistics. Upon alignment, structures were loaded into open-source PyMOL v3.1.0 for coloration and production of images of proteins, green for fused, blue for split, and red for basal forms of the proteins.

Results

Phylogenetic analysis reveals novel lichen-associated isocyanide synthase clades

We screened 3,992 fungal genomes for the isocyanide synthase (ICS) domain (PF05141.1) to find representative genomes for each species complex (see Methods). This filtering yielded a final

dataset of 1,847 ICS proteins from 814 genomes, including 39 from lichen-forming fungi (LFF) (Table S1, S2). To map the evolutionary diversity of isocyanide synthases (ICS) in lichen-forming fungi (LFF), we constructed a maximum-likelihood phylogeny from these 1,847 ICS proteins.

A minority of LFF-derived ICS proteins grouped with two previously characterized gene cluster families (GCFs) such as the *dit*-variant and GCF 4¹⁸ (Table S6). However, our analysis also uncovered six highly supported clades significantly enriched with LFF sequences (Fig. 1). We designated these "Lichen Group 1" (LG1) and "Lichen Group 2" (LG2), "Lichen Group 3" (LG3), "Lichen Group 4" (LG4), "putative *crmA*-type ICS", and "basal to *crmA*-type ICS". Given the exceptional metal tolerance of many lichens and the known role of *crmA* in metal ecology (*crm* = copper responsive metabolite), we focused our investigation on two of the novel lichen clades (LG1 and LG2) and the two LFF-rich clades phylogenetically associated with *crmA*. While LG3 and LG4 represent novel lineages, their detailed characterization was beyond the scope of this study and they remain promising subjects for future research.

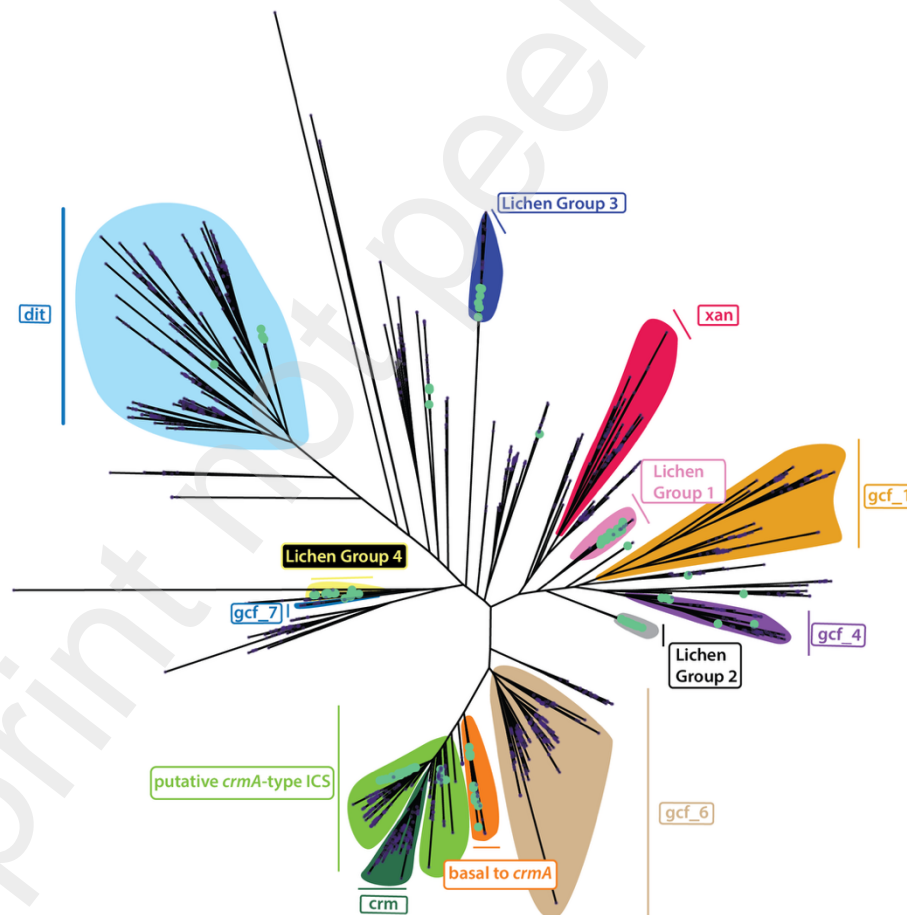


Figure 1. Unrooted phylogeny of fungal isocyanide synthases' (ICS) amino acid sequences. Clades corresponding to gene cluster families (GCFs) previously annotated by Nickles et al. (2003) are indicated on the tree. Six previously unreported, lichen-associated ICS clades investigated in this study

are highlighted: "Lichen Group 1" (LG1; pink), "Lichen Group 2" (LG2; grey), "Lichen Group 3" (LG3; dark blue), "Lichen Group 4" (LG4; yellow) and "putative/basal *crmA*-type ICS" (indicated in lime green and orange). Tips representing ICS proteins derived from lichen genomes are colored green.

Genomic analysis reveals a conserved set of genes surrounding LG1 and LG2

We investigated if there was *in silico* evidence for LG1 and LG2 being a part of BGCs or if they instead appeared to function as standalone ICS enzymes. The LG1 ICS, found in 27 species, is located within a highly conserved locus containing numerous BGC-associated genes, including a P450, multiple oxidoreductases, MFS transporters, and a transcription factor. This strong synteny provides clear evidence that LG1 is the backbone of a bona fide, lichen-associated BGC (Fig. 2).

LG2, in contrast to LG1, was found in a much smaller number of species: nine *Xylographa* and a single genome of *Lambiella insularis* (Fig. 2). While a core set of six genes is conserved between *Lambiella* and *Xylographa* species, the locus notably contains a putative primary metabolism enzyme, carbamoyl phosphate synthetase (CPSase), and an FAD binding hydroxylase enzyme alongside the LG2 ICS. (Fig. 2).

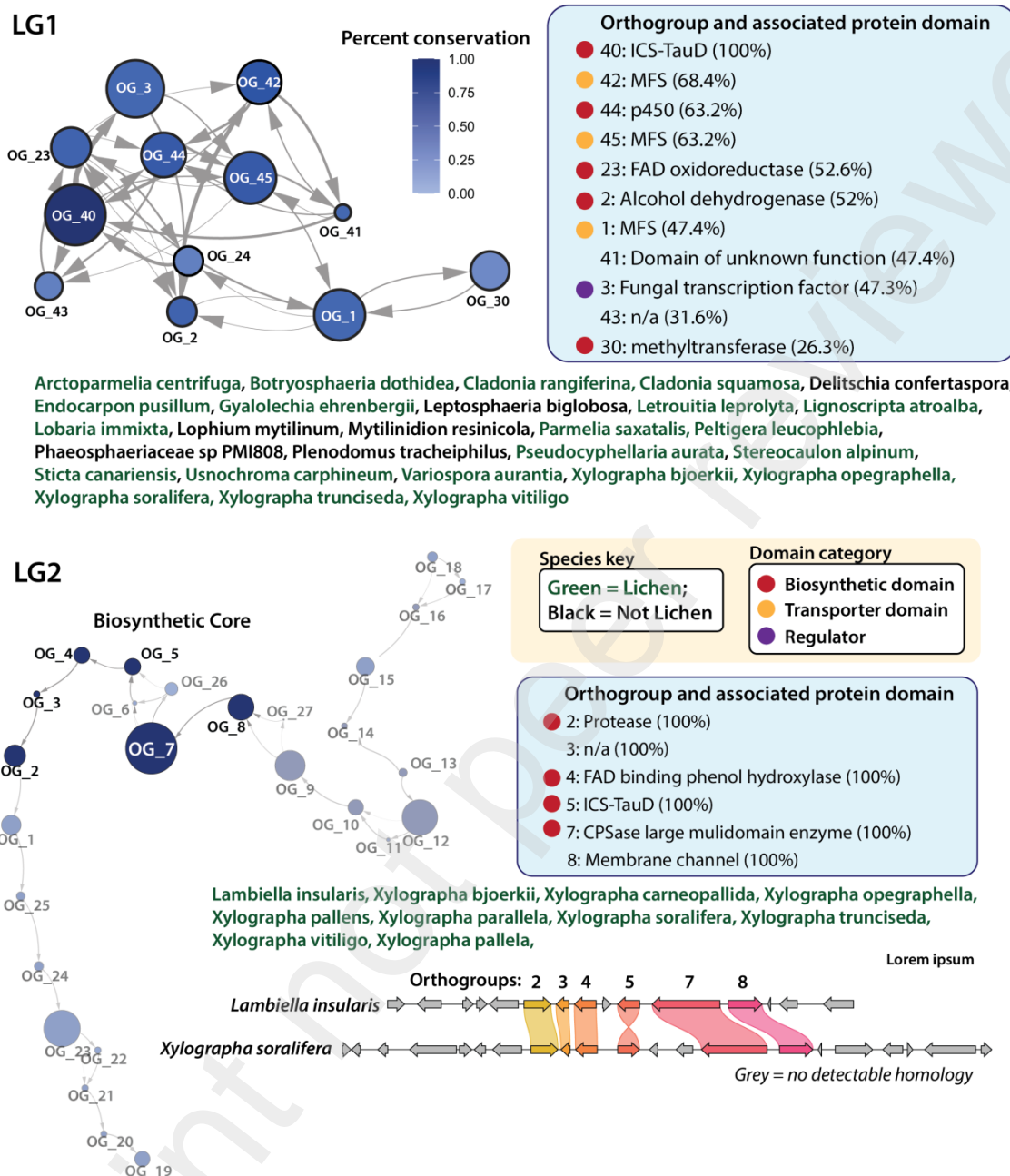


Figure 2. Domain structure of the loci surrounding lichen group 1 (LG1) and the lichen group 2 (LG2) isocyanide synthases (ICSs). Network graphs display orthogroups as nodes with edges representing gene order. Node size represents the median length of genes within that orthogroup. The protein domains in each orthogroup are detailed in the associated key tables (blue) for each network. The species containing LG1 and LG2 are listed; species names are colored green (lichen) or black (non-lichen) according to their ecology. For LG2, a locus alignment was added to illustrate the orthogroups specifically conserved between *Lambiella* and *Xylographa*.

Investigation of the putative *crmA*-type ICS reveals a split *crmA* architecture that is functionally equivalent to its canonical fused counterpart

Given the established role of the *crm* cluster in metal homeostasis, we focused our analysis on three distinct subclades predominantly comprising LFF (Lecanoromycetes) that were sister to the *crm* ICSs; collectively called putative or basal *crmA*-type ICS and comprising a total of 394 proteins (Fig. 1). Manual inspection of these loci surrounding the LFF ICS proteins in the putative or basal *crmA*-type ICS clades of proteins revealed that the ICS-encoding gene was frequently located next to a separate gene, often in opposite orientation, encoding the CrmA-signature NRPS-like domains of an adenylation (A), a peptidyl carrier protein (T), and a non-canonical transferase domain. Such split NRPS proteins shared greater than 40% amino acid similarity to the canonical fused domains in the *A. fumigatus* CrmA protein. This observation led us to hypothesize that these loci represented a "split" variant of the *crmA* pathway, where the ICS and NRPS-like components are encoded by separate genes.

A systematic analysis of the signature A-T-Transferase *crmA*-NRPS-like domains and their proximity to the phylogenetically identified *crmA*-ICS enzymes confirmed that the "split" architecture is widespread. Our pipeline located the signature NRPS-like gene near 350 of the 394 *crmA*-like ICS genes analyzed (88.83%). The corresponding NRPS-like gene was annotated as fused to the *crmA* gene in 142/394 (36.04%) of loci and split from the NRPS-like gene in 208/394 (52.79%) of loci. For the remaining 11.17% of loci our pipeline was unable to locate the NRPS-like gene on the same contig as the monodomain-*crmA* gene. The majority of split copies were encoded on opposite strands (81.25%). Many of the split ICS and NRPS encoding genes share an intergenic promoter region of 1-2kb (Fig. 4).

All *crmA* homologs identified within the Lecanoromycetes LFF genomes were of the split architecture. To test if the split *crmA* pathway is functional, we characterized the transcriptionally active split locus (FGSG_03735/FGSG_03736) in the agriculturally relevant and genetically tractable, *Fusarium graminearum* PH-1 (see Methods and Sup Methods). We were able to detect the known intermediate, N-formyl-L-valine, and final product fumicicolin A, of the fused-*crmA* pathway in our *F. graminearum* PH-1 strain, the genome of which only encodes split *crmA* genes (Fig. S1). Deletion of the split ICS gene FGSG_03735 (Δ ICS) abolished production of both the N-formyl-L-valine intermediate and the final product, fumicicolin A (Fig. 3a). In contrast, deletion of the split NRPS-like gene FGSG_03736 (Δ NRPS) abolished fumicicolin A production but led to the accumulation of the N-formyl-L-valine intermediate. Our genomic and metabolic knockout data confirm a functional split *crmA* pathway and shows that both the ICS and NRPS-like enzymes are required for fumicicolin A biosynthesis. This result aligns with a recent study by Kishimoto et al. (2025)²⁵, which showed that heterologous expression of a single fused *crmA* gene is sufficient for pinocicolin biosynthesis (Fig. 3b). Taken together, these findings provide strong evidence that the

Previously, the *crmA* pathway was conservatively defined by the presence of a multi-gene BGC homologous to that in the reference species, *A. fumigatus*^{18,23}. Redefining the *crmA* pathway to include both split and fused variants increases the number of non-Saccharomycotina Ascomycetes species likely producing *crmA*-related metabolites from 51/938 (5.4%) to 314/938 (33.4%), making *crmA* (fused or split) the most common ICS pathway in the fungal kingdom (Fig. S2).

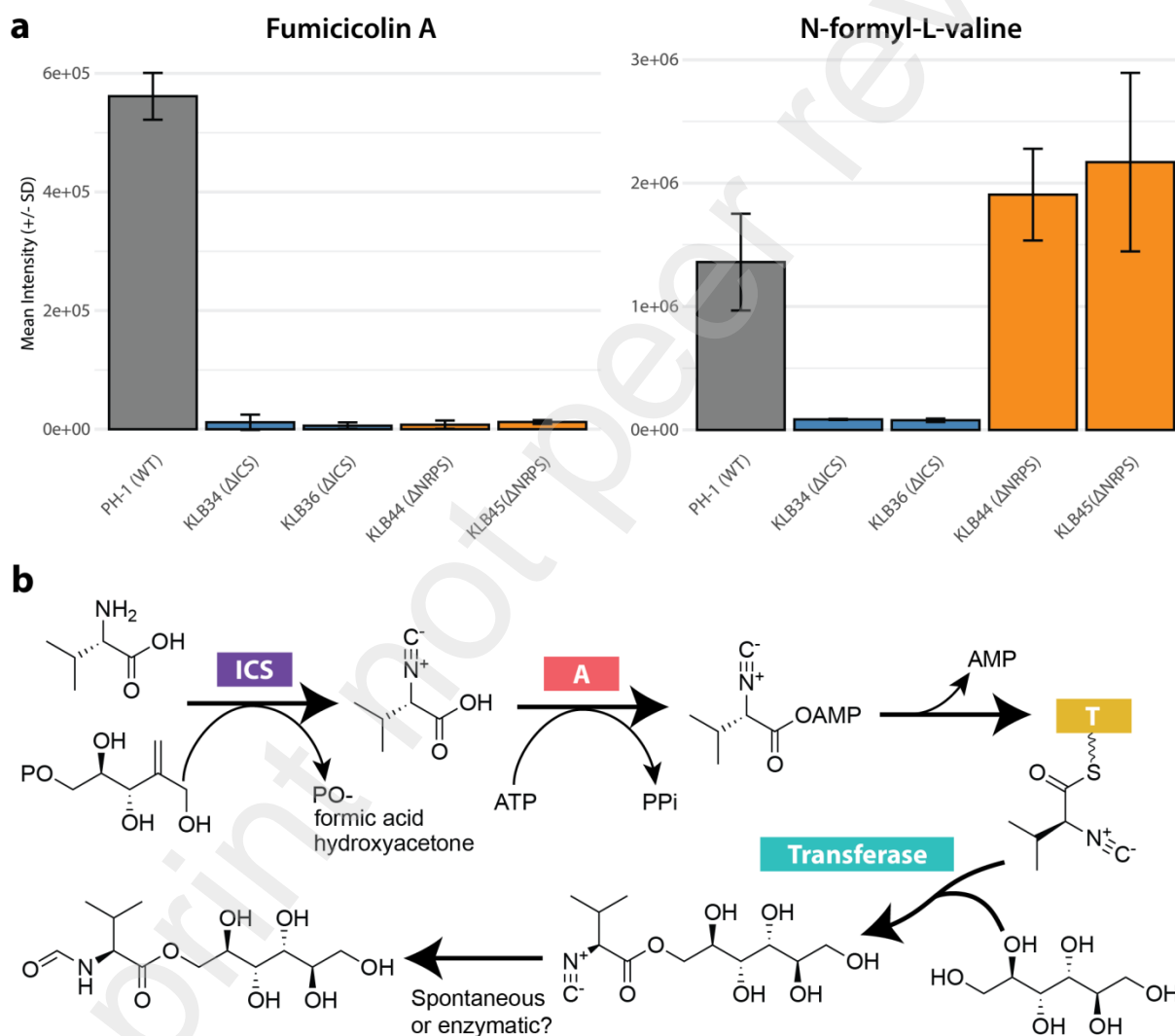


Figure 3. Metabolic characterization of putative fumicicolin A and N-formyl-L-valine in *Fusarium graminearum* PH-1 WT, $\Delta crmA$ -ICS, and $\Delta crmA$ -NRPS mutants. **(a)** Mean intensity values for the screened known MS/MS fragmentation patterns of fumicicolin A and N-formyl-L-valine. **(b)** Proposed fumicicolin A

biosynthetic pathway, based on the proposed pinocicolin A pathway from Kishimoto et al. (2025)²⁵ (recreated with permission).

The CrmA-associated NRPS-like protein exhibits diverse domain architectures

We identified ten distinct domain architectures for the CrmA-associated NRPSs, namely Type A through F (Fig 4), along with rare variants designated as Type G-J (Fig. S4). Approximately three-quarters (73.3%) of the identified NRPS-like proteins, regardless of whether they originated from a split or fused *crmA* locus, had the canonical A-T-transferase domain architecture (defined here as Type A; Fig. 4; Fig. S3). Notably, the fused Type A domain structure was the one reported in our previous study (Nickles et al. (2003); Table S7) and is predominately comprised of non-lichen forming fungi.

The majority of the other variants likely resulted from domain loss or truncation, rather than representing entirely novel structures. Type B proteins ($n = 37$) resembled Type A but lacked a detectable C-terminal transferase domain (due to mutation or absence). Other minor variants included Type F proteins ($n = 6$), which lacked a detectable A domain, and the single Type I protein, which lacked a detectable T domain (Fig. S4). Type E ($n = 6$) represented a significantly truncated protein containing only the A domain, potentially resulting from gene fragmentation or annotation errors.

Notably, the Type B architecture (mutated transferase) was predominantly found in *Fusarium* species, which often possess two paralogous *crmA* pathways (Fig. 4). This Type B architecture is the same as the transcriptionally silent NRPS-like paralog from PH-1 that we hypothesized might be non-functional (see Sup Methods). Within the *Fusarium* genus, if a species has *crmA*, one of the copies is always Type A. The other often will be Type B, although we did find a small number of *Fusarium* isolates with two Type A NRPS-like genes.

The Type D architecture matches Type A but has an additional ARN family metal transporter domain downstream of the transferase. Sixteen of the seventeen species with a Type D structure were in the fused clade, but there was a single Type D split copy NRPS detected with the genome of *Colletotrichum camelliae* (KAH0423461.1). If this domain architecture is not a product of annotation errors, the taxonomic distribution of this domain would require that either there were multiple independent fusion events of an ARN family metal transporter domain onto the *crmA* gene or that Type A arose from multiple ancient deletion events of a previous Type D ancestor.

Of particular interest was the Type C NRPS-like domain architecture, which was found exclusively in a basal clade of taxonomically diverse fungi (Fig. 1a, orange clade) and uniquely contained a canonical NRPS C-terminus thioester reductase (TR) domain instead of the Type A signature transferase. In the phylogeny of *crmA*-ICS domains that was rooted using the a fungal ICS-TauD domain, Type C proteins formed a basal paraphyletic group relative to (and ancestral to) the split

crmA clade (Fig. 3). In the rooted phylogeny based on the NRPS-like domain, Type C formed a basal monophyletic group, a discrepancy potentially explained by the influence of the TR domain sequence on the NRPS-like alignment and resulting tree topology (Fig S3).

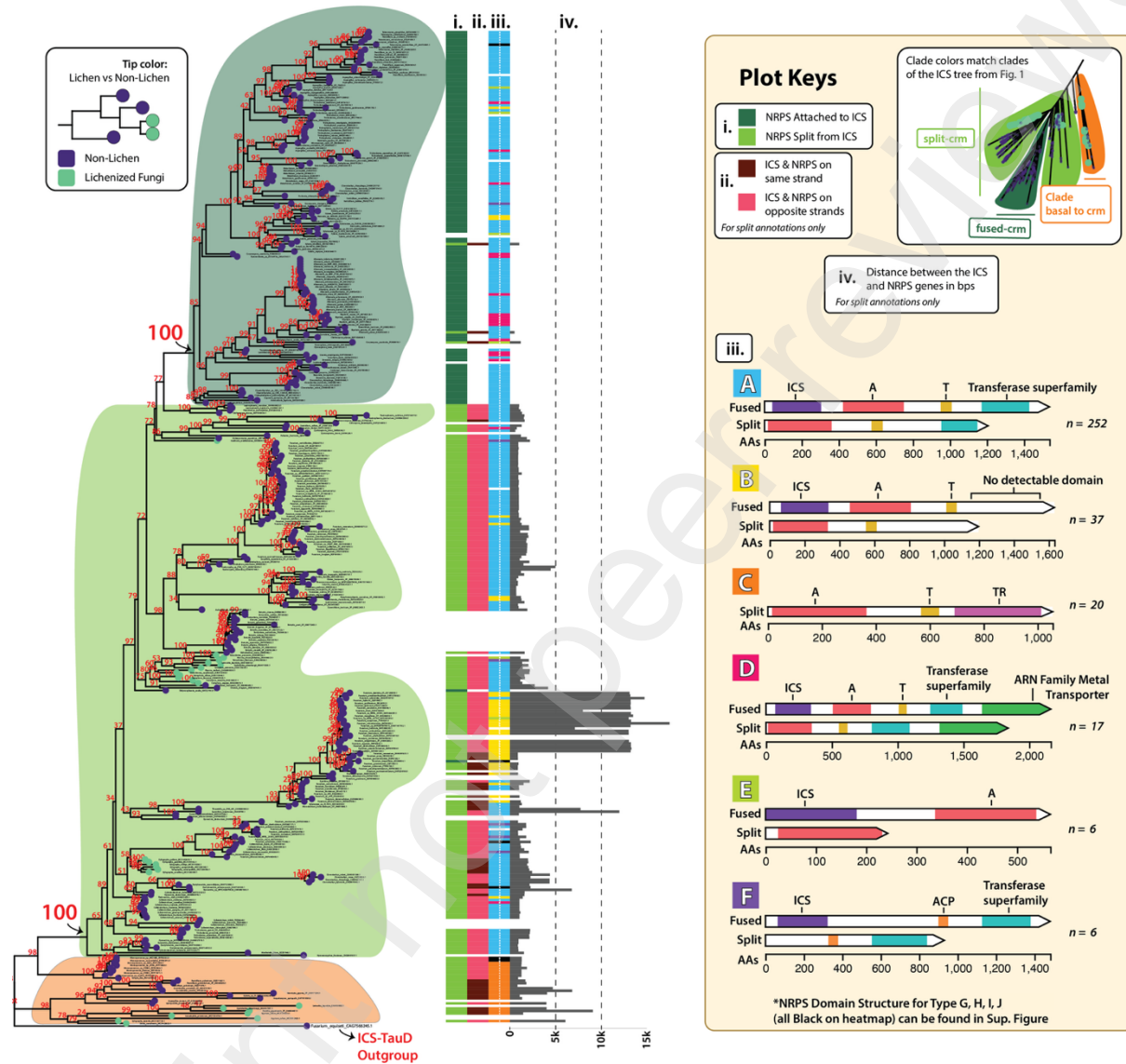


Figure 4. Reconstructed maximum-likelihood phylogeny of the newly defined *crmA*-type ICS clade. The tree is rooted using a known ICS-TauD protein sequence as the outgroup. Bootstrap values are displayed with red text. Associated heatmaps contain (from left to right): i) whether the NRPS-like region is fused to the ICS gene, ii) the relative orientation of the NRPS-like gene to the ICS gene, iii) the domain architecture of the NRPS-like proteins, and iv) the distance (in bp) between the ICS and separate NRPS-like genes (where applicable). Less common domain architectures of NRPS-like proteins (labeled G-J), which were found in only one (H, I, J) or two (G) distinct protein sequences in our dataset, can be found in Fig. S4.

Phylogenetic and structural data reveal the stepwise evolution of the fused CrmA megasynthase

To understand the origin of the fused *crmA* architecture, we compared the phylogenies of the ICS and NRPS-like protein domains encoded by all *crmA* loci (Fig. S5). Both the ICS and separate NRPS-like domain phylogenies independently resolved a single monophyletic group of fused proteins nested within a paraphyletic group of their split counterparts (Fig. 4, S2). The strong congruence between the two domain phylogenies (Kuhner-Felsenstein distance = 4.525) and the bootstrap support for these branches suggests there was a single fusion event from an ancestral split-gene pair.

We next investigated the evolution of the NRPS-like protein itself. The basal position of the Type C architecture suggests it is the ancestral form. This proposed evolutionary trajectory is supported by genomic and structural data. While the Type A domain architecture is found exclusively in *crmA* loci (Table S7), the ancestral Type C (A-T-TR) architecture is common in other BGCs across the fungal kingdom. As an example, we located in *Aspergillus fumigatus* Af293 (which possesses only a fused, Type A *crmA* gene) a single non-*crm* uncharacterized BGC with a gene encoding an A-T-TR (Type C) protein (Afu3g02670). Furthermore, structural alignments show that the core ICS, A, and T domains are highly conserved between Type C, Type A-split, and Type A-fused proteins (TM-scores > 0.7). In contrast, the ancestral TR domain and transferase domains are structurally unrelated (TM-score = 0.1) (Fig. 5b, Table S8).

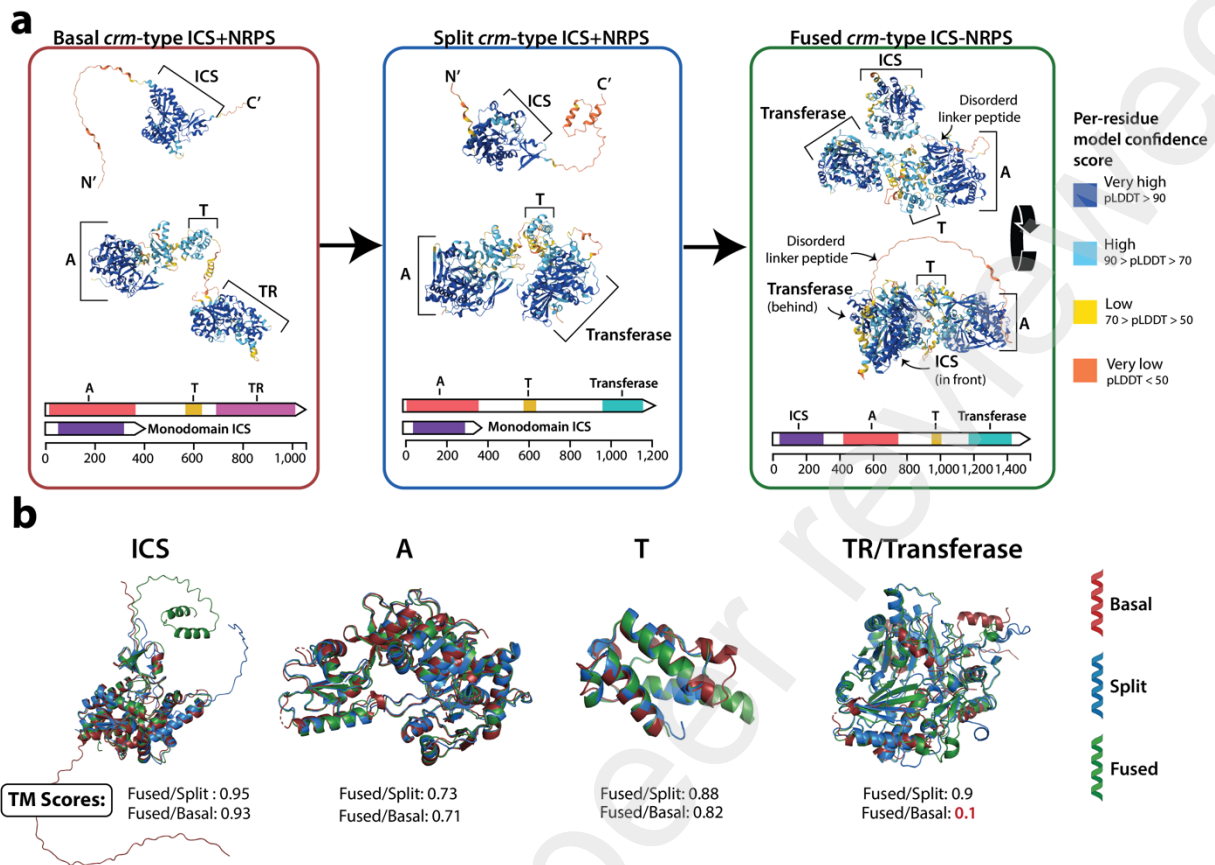


Figure 5. Proposed evolutionary model for the CrmA metabolic pathway, with supporting structural alignments of CrmA domains. **(a)** Protein structures are colored with the per-residue model confidence score. The evolutionary model depicts three stages. The proposed ancestral 'basal' state, Type C, consists of separate ICS and NRPS-like proteins and notably has a thioester reductase (TR) domain on the C-terminus. The split Type A architecture has a C-terminal transferase and is hypothesized to have evolved from Type C via full replacement of the TR domain for the transferase. Finally, the fused Type A architecture is hypothesized to have arisen via a gene fusion event within an ancestral split Type A lineage. **(b)** Pairwise structural alignments of individual CrmA domains (ICS, A, T, and the C-terminal TR or Transferase domains). Each domain was extracted from the full-length protein models shown in 'a' and aligned separately. TM-scores, quantifying structural similarity, are provided for each indicated pairwise alignment (e.g., 'fused/split' compares the domain from the fused CrmA with that from the split Type A form). The aligned structures are colored as follows: basal-red; split-blue; fused-green.

The CrmA phylogeny topology is incongruent to the species tree

Cophylogenetic analysis of the CrmA ICS-domain phylogeny and the fungal species tree revealed significant topological incongruence in both rooted and unrooted alignments (Fig. S6, S7). This was often characterized by entire taxonomic clades being misplaced in the gene tree relative to their expected position in the species tree. Notably, the internal branching patterns within these misplaced clades frequently remained congruent with known species relationships. We believe

these misplaced but internally ordered clades suggests the pathway's evolution has been shaped by extensive horizontal gene transfer (HGT) or differential loss of ancestral paralogs. Although the strong congruence between the ICS and NRPS-like phylogenies indicates the gene tree is robust, it remains possible that variable selective pressures among clades also contribute to the observed incongruence with the fungal species tree.

The transcriptional regulation of the split *crmA* pathway does not match the canonical fused *crmA* pathway

As previously stated, the fused *crmA* pathway has been investigated by our research group and others in *A. fumigatus*, *Alternaria brassicicola*, *Cochliobolus heterostrophus*, *Penicillium commune*, and *Penicillium pinophilum*^{21,23–25}. It has been shown in both *A. fumigatus* and *A. brassicicola* that the fused *crmA* gene and associated metabolites are sharply upregulated when the fungus is grown in copper starvation^{21,24}. We investigated if the split ICS (FGSG_03735) and NRPS (FGSG_03736) in *F. graminearum* PH-1 are similarly transcriptionally upregulated under copper starvation in two different minimal media made with and without copper (see Sup. Methods). We also checked the expression of the second *crmA* ICS gene (FGSG_08185), which was transcriptionally silent in all previously examined public datasets (see Sup. Methods). In contrast to the findings in *A. fumigatus* and *A. brassicicola*, the FGSG_03735 and FGSG_03736 genes had modest but significantly lower expression in the glucose and fusarium minimal media that lacked copper (Fig. S9). Interestingly, while mRNA for the second *crmA* ICS gene (FGSG_08185) was detected in our RT-qPCR experiment, its expression level did not differ between media types.

Discussion

Some lichens have the unique ability to grow in highly metal-contaminated areas due to the development of adaptive mechanisms, such as sequestering high metal concentrations in their thalli^{11,12}. The specialized metabolites produced by ICS BGCs have been linked to metal-associated pathogenicity and metal-sequestration in both bacteria²⁶ and fungi^{21,22,24}. As ICS BGCs were only canonically characterized in 2023¹⁸, and the number of publicly available Lecanoromycetes genomes has only recently grown to 50+, the association of ICSs within LFF was unknown. Given the established role of isocyanides in metal homeostasis and the high metal tolerance of many LFF, we reasoned that this fungal group could represent an untapped reservoir of ICS diversity. We reasoned that lichenized fungi were an ideal system to explore ICS diversity and evolution, given their frequent colonization of metal-rich environments and the established role of isocyanides in metal homeostasis.

Novel Isocyanide Synthase Lineages in Lichen-Forming Fungi

We found that lichen-forming fungi have four ICS clusters (LG1, 2, 3 and 4) that appear to be specific to them and are absent from other fungal groups. The first clade we investigated,

designated Lichen Group 1 (LG1), was found in 21 distinct lichenized Lecanoromycetes species as well as six non-lichen species. Co-localization of the same orthologous genes across diverse species and genera increases confidence that the genomic locus is preserved by selection. The conservation of many BGC-associated orthogroups in a wide range of species has previously been shown to be a strong indicator of a functional SM pathway conferring a selective advantage^{50,51}. Taken in conjunction with the presence of many genes encoding enzymes associated with secondary metabolism surrounding the LG1 loci (Fig. 2), we suggest this BGC is an excellent target for future heterologous expression. Furthermore, its potential role in metal ecology warrants investigation.

The second clade we investigated, LG2, was exclusively found in ten LFF Lecanoromycetes. Analysis of the genes surrounding LG2 revealed a set of six genes found in 100% of the ten species. Most interesting was the presence of a putative primary metabolism carbamoyl phosphate synthetase (CPSase) (Fig. 2). CPSases produce carbamoyl phosphate, which feeds into the arginine and pyrimidine biosynthesis pathways. The arginine pathway utilizes ornithine and produces citrulline, and arginine, while the pyrimidine biosynthesis pathway utilizes aspartate^{52,53}. If the CPSase and ICS are indeed working together, it is theoretically possible that one of the amino acids generated in these CPSase-related pathways could be a target for the LG2 ICS. We note that all characterized isocyanide synthases (ICSs) to date have been shown to exclusively convert the amino group on either tyrosine, valine, or tryptophan⁵⁴. Interestingly, CPSases are targets of the fungal SM phomoxanthone A, suggesting that the producing fungus contains an endogenous protective mechanism⁵⁵. Furthermore, CPSases have been associated with copper resistance in *A. fumigatus*⁵⁵. These findings suggest that deeper investigation into a possible linkage between this CPSase and ICS biology is warranted.

Why Fuse? The Evolutionary Rationale for the *crmA* Megasyntase

Our investigation into the genomic diversity of lichen-forming fungi unexpectedly uncovered a widespread, ancestral 'split' architecture of the *crmA* pathway, where the ICS and NRPS-like components are encoded as two separate genes. While not exclusive to lichens, this discovery highlights the power of exploring understudied taxonomic groups to reshape our understanding of major biosynthetic pathways.

Phylogenetic reconstruction of both the ICS and NRPS-like domains strongly indicates that the fused *crmA* megasyntase arose once from a split ancestral lineage, rather than the split form arising from a gene fission event. This evolutionary trajectory, where functionally related, co-localized genes are fused into a single ORF, is a recurring theme in the evolution of metabolic pathways, thought to be driven by selection for improved regulation or catalytic efficiency⁵⁶. The genomic architecture of the ancestral split genes, often oriented head-to-head, suggests a plausible molecular path to fusion involving a domain swap followed by gene inversion and

deletion of the intervening sequence. This model is consistent with our structural predictions of a flexible linker joining the two formerly separate proteins (Fig. 5, Sup. Video 1).

Crucially, this evolutionary transition did not result in neofunctionalization. Our findings in *F. graminearum* demonstrate that the split and fused architectures produce the same metabolites. This raises a key question: if the chemical output is unchanged, what selective advantage drove the fixation of the fused *crmA* pathway within Ascomycota?

While more work is needed to definitively determine the precise selective pressures that fixed the fused-*crmA* pathway within Ascomycota, the principle of metabolic channeling offers a compelling explanation. We propose this would be particularly advantageous for the *crmA* pathway due to the reactive nature of its isocyanide intermediate⁵⁷. For instance, our group previously showed that valine isocyanide can react *in trans* with precursors from the ergot alkaloid pathway to produce hybrid alkaloids called fumivalines²¹. Fusing the ICS and NRPS-like domains would provide a mechanism to sequester this reactive intermediate, preventing such metabolic crosstalk and maximizing flux towards the intended product. This aligns with the 'Used-Fused Hypothesis', which posits that genes encoding interacting proteins or enzymes in the same pathway are more likely to undergo fusion events that are subsequently maintained by selection.

These theoretical advantages are supported by findings in synthetic protein engineering. For example, Kummer et al. (2021) compared the rate of a two-step enzymatic reaction where two proteins were left separate, linked by a neutral linker, or linked with a rationally designed linker⁵⁸. This group reported increased catalytic turnover frequencies of 140-fold for the neutrally linked proteins relative to the separated enzymes and an even more dramatic 500-fold for the rationally linked proteins optimized for the active-site proximity⁵⁸.

Considering these precedents, we speculate that the disordered peptide linker identified in the fused CrmA structure could provide the necessary flexibility for optimal orientation of the ICS and NRPS-like domains, thereby facilitating efficient substrate channeling of valine isocyanide produced by the ICS domain directly to the NRPS-like domains. This theoretically could give it a selective benefit by increasing the flux of the pathway while minimizing the diffusion or potential toxicity of valine isocyanide generated by the ICS domain⁵⁹. Interestingly, the strategy of synthetically linking enzymes coordinating together in a pathway has rarely been applied to increase the production of microbial SMs but is becoming an increasingly used approach in industrial protein engineering and synthetic biology^{56,58–60}. A study that quantitatively compares the metabolic flux of the canonical fused CrmA (from valine and D-mannitol to fumicolin A) with that of an engineered 'separated' version, where the ICS and NRPS-like domains are expressed as distinct, unlinked proteins could enable an empirical test to determine definitively if the fused architecture improves the metabolic flux of this highly unusual hybrid ICS-NRPS-like pathway, where we would predict a lower flux for the split version.

The widespread distribution of fused and split *crmA* pathways

The fused *crmA* gene has been shown to be sharply upregulated under copper starvation in *A. fumigatus*, and *A. brassicicola* who both contain the fused pathway^{21,24}. In contrast, the split *crmA* genes in *F. graminearum* were not induced by copper starvation but were instead modestly downregulated. This divergence suggests that integration of the *crmA* pathway into the copper-responsive network may be a derived trait that co-evolved with the fusion event.

The core *crmA* gene pair (ICS and associated NRPS-like, regardless of split or fused architecture) was detected in 33.4% of the non-Saccharomycotina Ascomycetes, or 29.48% of all Ascomycetes, in our database (Fig. S8, Fig. S2). This identifies the *crmA* pathway as the most common type of isocyanide specialized metabolism pathway in the fungal kingdom, and its prevalence rivals that of even the most foundational secondary metabolic pathways, such as those for DHN-melanin⁶¹ and siderophore biosynthesis⁶², which are considered nearly ubiquitous across the phylum. For context, it was previously found that less than 5% all canonical fungal gene cluster families are found in taxonomic ranks higher than family⁶³, yet the *crmA* split and fused pathway spanned fungal taxonomic Classes.

The cophylogenetic analysis presented in this study shows significant topological incongruence between the *crm* ICS-domain tree and the species tree. Future research should aim to more definitively distinguish the relative contributions of HGT versus differential loss of ancestral paralogs to the observed incongruent *crmA* topology (Fig. S6, S7). To this end, a focused investigation of major phylogenetically incongruent *crmA* clades should seek independent lines of genomic evidence typically associated with HGT, such as atypical codon usage patterns or divergent GC-content within the hypothesized transferred loci relative to their respective host genomes.

Resources availability

Requests for further information and resources should be directed to the two lead contacts, Dr. Nancy Keller (npkeller@wisc.edu) and Dr. Garima Singh (garima.singh@unipd.it). No unique reagents were generated for this publication, all genomes and RNA-seq datasets used were obtained from publicly available databases. A supplemental repository has been created to store all generated phylogenetic trees and all structural alignment files ([GitHub Link to Repository](#)). The supplemental information file hosted by Current Biology contains all supplementary figures, tables and methods. The supplemental information file is also stored in the supplemental repository.

Author Contributions

Conceptualization: GRN, AV, MTD, NPK, GS; Data curation: GRN, AV, GS; Formal analysis: GRN, AV, GS; Funding acquisition: GRN, MTD, NPK, GS; Investigation: GRN, AV, SCP, KB, HPE; Methodology:

GRN; Project administration: GRN, GS; Resources: MTD, NPK, GS; Software: GRN, AV, GS; Supervision: MTD, NPK, GS; Validation: GRN; Visualization: GRN; Writing – original draft: GRN, AV, SCP, HPE, GS; Writing – review & editing: GRN, AV, SCP, KB, HPE, MTD, NPK, GS

Declaration of interests

The authors declare no competing interests.

Supplemental information

Supplemental information can be found online at:

https://github.com/gnick18/ICS_Lecanoromycetes_crmA_SupRepo

Acknowledgments

G.S. was supported by the Italian Ministry of University and Research (project funded by the European Union—Next Generation EU: “PNRR Missione 4 Componente 2, “Dalla ricerca all’impresa”, Investimento 1.4, Progetto CN00000033”). N.P.K. was supported by National Institutes of Health R35 GM 156119 – 01; and H.P.E. by National Science Foundation Graduate Research Fellowship 2137424. G.R.N. was supported by National Science Foundation Graduate Research Fellowship [2137424 to G.R.N.] and National Institute of General Medical Sciences T32 [GM135066 to G.R.N.]. M.D. was supported by United States Department of Agriculture-Agricultural Research Service project 5062-21220-024-000D. Mention of trade names or commercial products in this article solely for the purpose of providing specific information and does not imply recommendation or endorsement by the United States Department of Agriculture (USDA). USDA is an equal opportunity provider and employer.

References

1. Hawksworth, D.L., and Grube, M. (2020). Lichens redefined as complex ecosystems. *N. Phytol.* 227, 1281–1283. <https://doi.org/10.1111/nph.16630>.
2. Allen, J.L., and Lendemer, J.C. (2022). A call to reconceptualize lichen symbioses. *Trends Ecol. Evol.* 37, 582–589. <https://doi.org/10.1016/j.tree.2022.03.004>.

3. Spribille, T., Thor, G., Bunnell, F.L., Goward, T., and Björk, C.R. (2008). Lichens on dead wood: species-substrate relationships in the epiphytic lichen floras of the Pacific Northwest and Fennoscandia. *Ecography* 31, 741–750. <https://doi.org/10.1111/j.1600-0587.2008.05503.x>.
4. Malek, J., Palice, Z., and Vondrk, J. (2014). New Lichen Records and Rediscoveries from the Czech Republic and Slovakia. *Herzogia* 27, 257–284. <https://doi.org/10.13158/heia.27.2.2014.257>.
5. Paukov, A., Teptina, A., Ermoshin, A., Kruglova, E., and Shabardina, L. (2022). The Role of Secondary Metabolites and Bark Chemistry in Shaping Diversity and Abundance of Epiphytic Lichens. *Front. For. Glob. Chang.* 5, 828211. <https://doi.org/10.3389/ffgc.2022.828211>.
6. Lutzoni, F., Kauff, F., Cox, C.J., McLaughlin, D., Celio, G., Dentinger, B., Padamsee, M., Hibbett, D., James, T.Y., Baloch, E., et al. (2004). Assembling the fungal tree of life: progress, classification, and evolution of subcellular traits. *Am. J. Bot.* 91, 1446–1480. <https://doi.org/10.3732/ajb.91.10.1446>.
7. Hibbett, D.S., Binder, M., Bischoff, J.F., Blackwell, M., Cannon, P.F., Eriksson, O.E., Huhndorf, S., James, T., Kirk, P.M., Lücking, R., et al. (2007). A higher-level phylogenetic classification of the Fungi. *Mycol. Res.* 111, 509–547. <https://doi.org/10.1016/j.mycres.2007.03.004>.
8. Schoch, C.L., Sung, G.-H., López-Giráldez, F., Townsend, J.P., Miadlikowska, J., Hofstetter, V., Robbertse, B., Matheny, P.B., Kauff, F., Wang, Z., et al. (2009). The Ascomycota Tree of Life: A Phylum-wide Phylogeny Clarifies the Origin and Evolution of Fundamental Reproductive and Ecological Traits. *Syst. Biol.* 58, 224–239. <https://doi.org/10.1093/sysbio/syp020>.
9. Gueidan, C., Hill, D.J., Miadlikowska, J., and Lutzoni, F. (2015). Systematics and Evolution. 89–120. https://doi.org/10.1007/978-3-662-46011-5_4.
10. Díaz-Escandón, D., Tagirdzhanova, G., Vanderpool, D., Allen, C.C.G., Aptroot, A., Češka, O., Hawksworth, D.L., Huereca, A., Knudsen, K., Kocourková, J., et al. (2022). Genome-level analyses resolve an ancient lineage of symbiotic ascomycetes. *Curr. Biol.* 32, 5209–5218.e5. <https://doi.org/10.1016/j.cub.2022.11.014>.
11. Rola, K. (2020). Insight into the pattern of heavy-metal accumulation in lichen thalli. *J. Trace Elem. Med. Biol.* 61, 126512. <https://doi.org/10.1016/j.jtemb.2020.126512>.
12. Rola, K., Osyczka, P., and Kafel, A. (2016). Different Heavy Metal Accumulation Strategies of Epilithic Lichens Colonising Artificial Post-Smelting Wastes. *Arch. Environ. Contam. Toxicol.* 70, 418–428. <https://doi.org/10.1007/s00244-015-0180-5>.
13. Silva, D.D.D., Rapior, S., Sudarman, E., Stadler, M., Xu, J., Alias, S.A., and Hyde, K.D. (2013). Bioactive metabolites from macrofungi: ethnopharmacology, biological activities and chemistry. *Fungal Divers.* 62, 1–40. <https://doi.org/10.1007/s13225-013-0265-2>.
14. Keller, N.P. (2019). Fungal secondary metabolism: regulation, function and drug discovery. *Nat Rev Microbiol* 17, 167–180. <https://doi.org/10.1038/s41579-018-0121-1>.

15. Gill, H., Sykes, E.M.E., Kumar, A., and Sorensen, J.L. (2023). Isolation of Bioactive Metabolites from Soil Derived Fungus-*Aspergillus fumigatus*. *Microorganisms* 11, 590. <https://doi.org/10.3390/microorganisms11030590>.
16. Bertrand, C., Prigent-Combaret, C., and Gonzales-Coloma, A. (2018). Chemistry, activity, and impact of plant biocontrol products. *Environ. Sci. Pollut. Res.* 25, 29773–29774. <https://doi.org/10.1007/s11356-018-3209-2>.
17. Singh, G., Calchera, A., Merges, D., Valim, H., Otte, J., Schmitt, I., and Grande, F.D. (2022). A Candidate Gene Cluster for the Bioactive Natural Product Gyrophoric Acid in Lichen-Forming Fungi. *Microbiol. Spectr.* 10, e00109-22. <https://doi.org/10.1128/spectrum.00109-22>.
18. Nickles, G.R., Oestereich, B., Keller, N.P., and Drott, M.T. (2023). Mining for a new class of fungal natural products: the evolution, diversity, and distribution of isocyanide synthase biosynthetic gene clusters. *Nucleic Acids Res.* 51, 7220–7235. <https://doi.org/10.1093/nar/gkad573>.
19. Ureña-Vacas, I., González-Burgos, E., Divakar, P.K., and Gómez-Serranillos, M.P. (2023). Lichen Depsides and Tridepsides: Progress in Pharmacological Approaches. *J. Fungi* 9, 116. <https://doi.org/10.3390/jof9010116>.
20. Asif, A., Iftikhar, A., Hamood, A., Colmer-Hamood, J.A., and Qaisar, U. (2019). Isonitrile-functionalized tyrosine modulates swarming motility and quorum sensing in *Pseudomonas aeruginosa*. *Microb. Pathog.* 127, 288–295. <https://doi.org/10.1016/j.micpath.2018.12.001>.
21. Won, T.H., Bok, J.W., Nadig, N., Venkatesh, N., Nickles, G., Greco, C., Lim, F.Y., González, J.B., Turgeon, B.G., Keller, N.P., et al. (2022). Copper starvation induces antimicrobial isocyanide integrated into two distinct biosynthetic pathways in fungi. *Nat Commun* 13, 4828. <https://doi.org/10.1038/s41467-022-32394-x>.
22. Raffa, N., Won, T.H., Sukowaty, A., Candor, K., Cui, C., Halder, S., Dai, M., Landero-Figueroa, J.A., Schroeder, F.C., and Keller, N.P. (2021). Dual-purpose isocyanides produced by *Aspergillus fumigatus* contribute to cellular copper sufficiency and exhibit antimicrobial activity. *Proc. Natl. Acad. Sci.* 118, e2015224118. <https://doi.org/10.1073/pnas.2015224118>.
23. Lim, F.Y., Won, T.H., Raffa, N., Baccile, J.A., Wisecaver, J., Keller, N.P., Rokas, A., and Schroeder, F.C. (2018). Fungal isocyanide synthases and xanthocillin biosynthesis in *Aspergillus fumigatus*. *Mbio* 9. <https://doi.org/10.1128/mbio.00785-18>.
24. Nadig, N., Park, S.C., Bok, J.W., and Keller, N.P. (2023). Conserved copper regulation of the antimicrobial isocyanide brassicicolin A in *Alternaria brassicicola*. *Fungal Genet. Biol.* 169, 103839. <https://doi.org/10.1016/j.fgb.2023.103839>.
25. Kishimoto, S., Takahashi, R., Nagasawa, A., and Watanabe, K. (2025). Discovery, Biological Activity, and Biosynthesis of Pinocicolin A, an Antibiotic Isocyanide Metabolite Produced by *Penicillium pinophilum*. *J. Nat. Prod.* <https://doi.org/10.1021/acs.jnatprod.5c00174>.

26. Crawford, J.M., Portmann, C., Zhang, X., Roeffaers, M.B.J., and Clardy, J. (2012). Small molecule perimeter defense in entomopathogenic bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 109, 10821–10826. <https://doi.org/10.1073/pnas.1201160109>.
27. Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., et al. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25, 1422–1423. <https://doi.org/10.1093/bioinformatics/btp163>.
28. Potter, S.C., Luciani, A., Eddy, S.R., Park, Y., Lopez, R., and Finn, R.D. (2018). HMMER web server: 2018 update. *Nucleic Acids Research* 46, W200–W204. <https://doi.org/10.1093/nar/gky448>.
29. El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Research* 47, D427–D432. <https://doi.org/10.1093/nar/gky995>.
30. Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., and Zdobnov, E.M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution*, 1–8. <https://doi.org/10.1093/molbev/msab199>.
31. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* 30, 772–780. <https://doi.org/10.1093/molbev/mst010>.
32. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>.
33. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., Haeseler, A. von, and Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* 37, 1530–1534. <https://doi.org/10.1093/molbev/msaa015>.
34. Hoang, D.T., Chernomor, O., Haeseler, A. von, Minh, B.Q., and Vinh, L.S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* 35, 518–522. <https://doi.org/10.1093/molbev/msx281>.
35. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., Haeseler, A.V., and Jermiin, L.S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods* 14, 587–589. <https://doi.org/10.1038/nmeth.4285>.
36. Letunic, I., and Bork, P. (2024). Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* 52, W78–W82. <https://doi.org/10.1093/nar/gkae268>.

37. Yu, G., Smith, D.K., Zhu, H., Guan, Y., and Lam, T.T.Y. (2017). Ggtree: an R Package for Visualization and Annotation of Phylogenetic Trees With Their Covariates and Other Associated Data. *Methods in Ecology and Evolution* 8, 28–36. <https://doi.org/10.1111/2041-210x.12628>.
38. Salamzade, R., Tran, P.Q., Martin, C., Manson, A.L., Gilmore, M.S., Earl, A.M., Anantharaman, K., and Kalan, L.R. (2025). zol and fai: large-scale targeted detection and evolutionary investigation of gene clusters. *Nucleic Acids Res.* 53, gkaf045. <https://doi.org/10.1093/nar/gkaf045>.
39. Blin, K., Shaw, S., Kloosterman, A.M., Charlop-Powers, Z., van Wezel, G.P., Medema, M.H., and Weber, T. (2021). antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* 49, gkab335-. <https://doi.org/10.1093/nar/gkab335>.
40. Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., and Warnow, T. (2014). ASTRAL: Genome-scale coalescent-based species tree estimation. *Bioinformatics* 30, 541–548. <https://doi.org/10.1093/bioinformatics/btu462>.
41. Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19, 15–30. <https://doi.org/10.1186/s12859-018-2129-y>.
42. Li, Y., Steenwyk, J.L., Chang, Y., Wang, Y., James, T.Y., Stajich, J.E., Spatafora, J.W., Groenewald, M., Dunn, C.W., Hittinger, C.T., et al. (2021). A genome-scale phylogeny of the kingdom Fungi. *Current Biology* 31, 1653-1665.e5. <https://doi.org/10.1016/j.cub.2021.01.074>.
43. Gilchrist, C.L.M., and Chooi, Y.-H. (2021). Syntheser: a CD-Search enabled Python toolkit for analysing domain architecture of fungal secondary metabolite megasynth(et)ases. *Fungal Biol. Biotechnol.* 8, 13. <https://doi.org/10.1186/s40694-021-00120-9>.
44. Consortium, T.U., Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Adesina, A., Ahmad, S., Bowler-Barnett, E.H., Bye-A-Jee, H., Carpentier, D., et al. (2024). UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic Acids Res.* 53, D609–D617. <https://doi.org/10.1093/nar/gkae1010>.
45. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. <https://doi.org/10.1038/s41586-021-03819-2>.
46. Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., et al. (2021). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 50, D439–D444. <https://doi.org/10.1093/nar/gkab1061>.
47. Varadi, M., Bertoni, D., Magana, P., Paramval, U., Pidruchna, I., Radhakrishnan, M., Tsenkov, M., Nair, S., Mirdita, M., Yeo, J., et al. (2024). AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res.* 52, D368–D375. <https://doi.org/10.1093/nar/gkad1011>.

48. Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>.
49. Bittrich, S., Segura, J., Duarte, J.M., Burley, S.K., and Rose, Y. (2024). RCSB protein Data Bank: exploring protein 3D similarities via comprehensive structural alignments. *Bioinformatics* 40, btae370. <https://doi.org/10.1093/bioinformatics/btae370>.
50. Cimermancic, P., Medema, M.H., Claesen, J., Kurita, K., Brown, L.C.W., Mavrommatis, K., Pati, A., Godfrey, P.A., Koehrsen, M., Clardy, J., et al. (2014). Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* 158, 412–421. <https://doi.org/10.1016/j.cell.2014.06.034>.
51. Chase, A.B., Sweeney, D., Muskat, M.N., Guillén-Matus, D.G., and Jensen, P.R. (2021). Vertical Inheritance Facilitates Interspecies Diversification in Biosynthetic Gene Clusters and Specialized Metabolites. *mBio* 12, e02700–21. <https://doi.org/10.1128/mbio.02700-21>.
52. Holden, H.M., Thoden, J.B., and Raushel, F.M. (1999). Carbamoyl phosphate synthetase: an amazing biochemical odyssey from substrate to product. *Cell. Mol. Life Sci. CMLS* 56, 507–522. <https://doi.org/10.1007/s000180050448>.
53. Shi, D., Caldovic, L., and Tuchman, M. (2018). Sources and Fates of Carbamyl Phosphate: A Labile Energy-Rich Molecule with Multiple Facets. *Biology* 7, 34. <https://doi.org/10.3390/biology7020034>.
54. Chen, T.Y., Chen, J., Tang, Y., Zhou, J., Guo, Y., and Chang, W. chen (2021). Current Understanding toward Isonitrile Group Biosynthesis and Mechanism. *Chinese Journal of Chemistry* 39, 463–472. <https://doi.org/10.1002/cjoc.202000448>.
55. Ceccacci, S., Deitersen, J., Mozzicafreddo, M., Morretta, E., Proksch, P., Wesselborg, S., Stork, B., and Monti, M.C. (2020). Carbamoyl-Phosphate Synthase 1 as a Novel Target of Phomoxanthone A, a Bioactive Fungal Metabolite. *Biomolecules* 10, 846. <https://doi.org/10.3390/biom10060846>.
56. Bolotin, E., Melamed, D., and Livnat, A. (2023). Genes that are Used Together are More Likely to be Fused Together in Evolution by Mutational Mechanisms: A Bioinformatic Test of the Used-Fused Hypothesis. *Evol. Biol.* 50, 30–55. <https://doi.org/10.1007/s11692-022-09579-9>.
57. Galli, U., Tron, G.C., Purgè, B., Grosa, G., and Aprile, S. (2020). Metabolic Fate of the Isocyanide Moiety: Are Isocyanides Pharmacophore Groups Neglected by Medicinal Chemists? *Chemical Research in Toxicology* 33, 955–966. <https://doi.org/10.1021/acs.chemrestox.9b00504>.
58. Kummer, M.J., Lee, Y.S., Yuan, M., Alkotaini, B., Zhao, J., Blumenthal, E., and Minter, S.D. (2021). Substrate Channeling by a Rationally Designed Fusion Protein in a Biocatalytic Cascade. *JACS Au* 1, 1187–1197. <https://doi.org/10.1021/jacsau.1c00180>.

59. Pareek, V., Sha, Z., He, J., Wingreen, N.S., and Benkovic, S.J. (2021). Metabolic channeling: predictions, deductions, and evidence. *Mol. Cell* 81, 3775–3785. <https://doi.org/10.1016/j.molcel.2021.08.030>.
60. Fu, J., Yang, Y.R., Johnson-Buck, A., Liu, M., Liu, Y., Walter, N.G., Woodbury, N.W., and Yan, H. (2014). Multi-enzyme complexes on DNA scaffolds capable of substrate channelling with an artificial swinging arm. *Nat. Nanotechnol.* 9, 531–536. <https://doi.org/10.1038/nnano.2014.100>.
61. Chamilos, G., and Carvalho, A. (2012). *Aspergillus fumigatus* DHN-Melanin (Springer International Publishing) <https://doi.org/10.1007/82>.
62. Haas, H. (2014). Fungal siderophore metabolism with a focus on *Aspergillus fumigatus*. *Natural Product Reports* 31, 1266–1276. <https://doi.org/10.1039/c4np00071d>.
63. Robey, M.T., Caesar, L.K., Drott, M.T., Keller, N.P., and Kelleher, N.L. (2021). An Interpreted Atlas of Biosynthetic Gene Clusters from 1000 Fungal Genomes. *PNAS* 118, <https://doi.org/10.1073/pnas.2020230118>.
64. Hall, C., Brachat, S., and Dietrich, F.S. (2005). Contribution of Horizontal Gene Transfer to the Evolution of *Saccharomyces cerevisiae*. *Eukaryot. Cell* 4, 1102–1115. <https://doi.org/10.1128/ec.4.6.1102-1115.2005>.
65. Milner, D.S., Attah, V., Cook, E., Maguire, F., Savory, F.R., Morrison, M., Müller, C.A., Foster, P.G., Talbot, N.J., Leonard, G., et al. (2019). Environment-dependent fitness gains can be driven by horizontal gene transfer of transporter-encoding genes. *Proc. Natl. Acad. Sci.* 116, 5613–5622. <https://doi.org/10.1073/pnas.1815994116>.