EXPANDING FUNCTIONAL ACCESS TO FUNGAL NATURAL PRODUCTS WITH METABOLOGENOMICS AND HETEROLOGOUS EXPRESSION

Benjamin Knox[1], Fatma Butun[2], Matthew Robey[2], Navid Ayon[2], Raveena Gupta[2], David Dainko[2], Jin Woo Bok[3], Grant Nickles[3], Robb Stankey[1], Don Johnson[1], David Mead[1], Kristof Cank[4], Cody Earp[4], Huzefa Raja[4], Nicholas Oberlies[4], Nancy P. Keller[1,3], Neil Kelleher[1,2], and Lindsay Caesar[2]

(1) Terra Bioforge, Madison WI (2) Northwestern University, Evanston IL

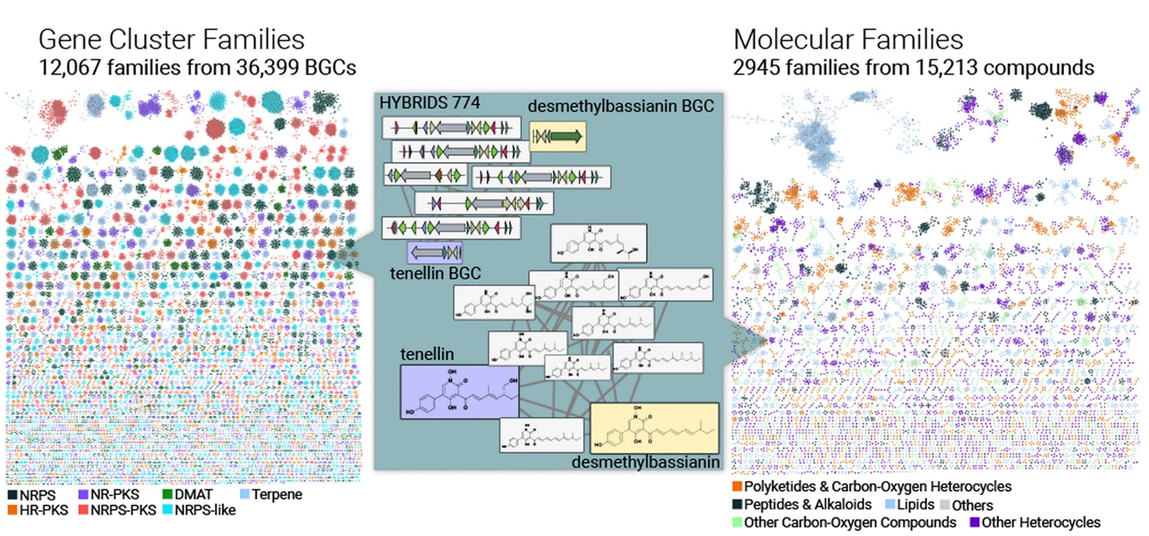
(3) University of Wisconsin-Madison, Madison WI (4) University of North Carolina, Greensboro NC

terra bioforge™

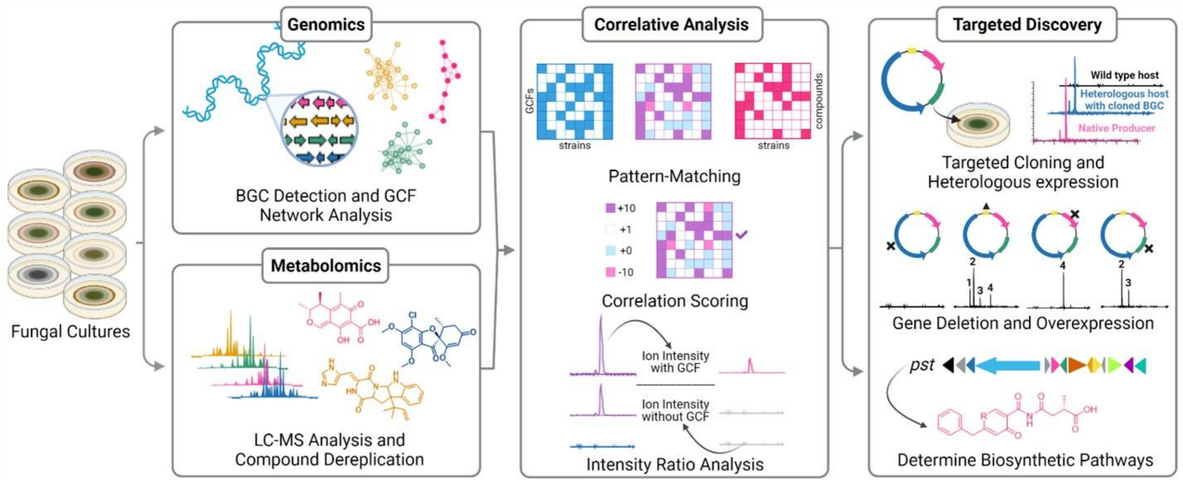
www.terrabioforge.com

ABSTRACT

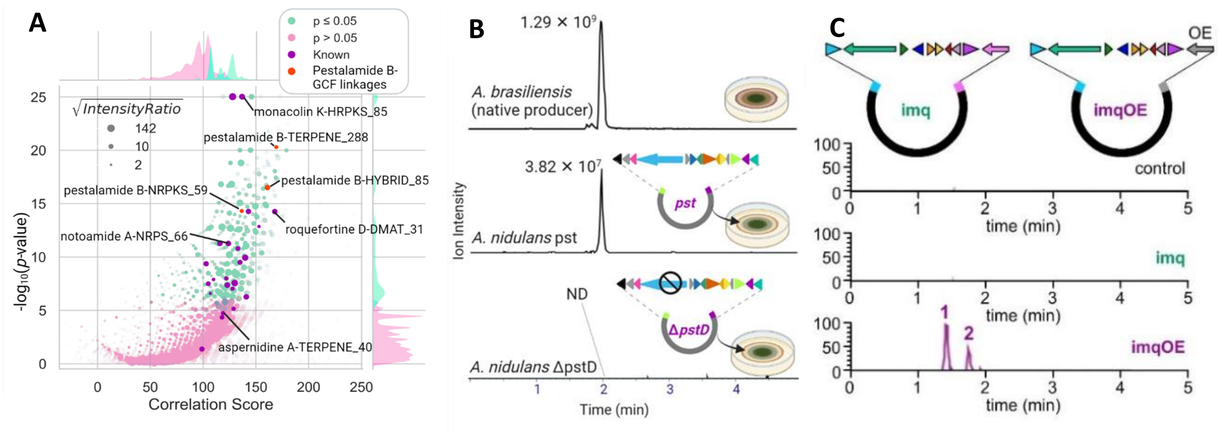
Fungal natural products (NPs) are valued for their unique and diverse bioactivities. As the fields' understanding of fungal NP chemical diversity continues to expand – driven by expansive genome sequencing efforts – limitations in current molecular biology methods have stymied rapid access to the molecules encoded by *in silico* predictions of biosynthetic gene clusters (BGCs). To address this, our group has conducted a global analysis of >1,000 public fungal genomes and annotated almost 12,000 gene cluster families that contain roughly 37,000 BGCs. To chemically annotate this dataset with actual molecules, we performed metabologenomics on 110 fungal strain extracts with MS-based metabolomics that link specific metabolite signatures to their co-occurring BGC families. Importantly, this analysis was validated with 25 known NP:BGC relationships. The outcome of this work is significant confidence in which BGCs to experimentally pursue for expressing a NP of interest and thousands of new NP correlations with their respective gene cluster families – a rich, mineable dataset for future discovery. Using CRISPR-based enzymes we precisely capture intact BGCs regardless of sequence content from complex gDNA to clone and refactor (as needed) into expression vectors. With this approach the pestalamide and imizoquin BGCs were transformed into an *Aspergillus nidulans* host optimized for heterologous expression – resulting in the production and detection of these NPs and proved the correct biosynthetic pathway of pestalamide. Ongoing activities within our group are to expand this fungal metabologenomics database to include NP extracts from hundreds of additional species, greatly expanding the pool from which we mine specific BGCs that are confidently linked to a natural product of interest. This approach is expected to facilitate the discovery of new molecules as well as chemical analogues of existing structures. Additionally, we are developing a brand-new approach for engineering fungi to achieve maximum heterologous expression titers all while maintaining high levels of genomic stability. Taken together, our group is establishing a new set of reliable tools to discover, produce, and functionally characterize fungal natural products.



***In silico* analysis pairs fungal gene cluster families (GCF) with molecular families (MF).** From 1,037 fungal genomes analyzed with antiSMASH (Blin *et al.* 2021) 36,399 biosynthetic gene clusters (BGC) were organized into 12,067 GCFs based on their putative biosynthetic similarities. 15,213 metabolites from the Natural Products Atlas (J. A. van Santen *et al.* 2019) were organized into MFs based on chemical relatedness. Anchoring GCFs with known BGCs allowed automated annotation of 2,026 BGCs with possible metabolite scaffolds which is roughly a tenfold increase in the number of annotated BGCs found in public databases. Importantly, the similar metabolites tenellin and desmethylbassianin and their respective BGCs were grouped in the same GCF and MF. 92% of the 12,067 GCFs are still unassigned to metabolite products – offering a significant opportunity to characterize new fungal natural products for pairing to their respective biosynthetic machinery.



Metabologenomics chemically annotates GCFs to streamline natural product (NP) characterization and production. Metabolomics data from 110 fungal strains each grown on three distinct medias was correlated with GCFs to pair NPs with their respective GCFs. Targeted cloning with CRISPR-based enzymes and heterologous expression offers a rapid method to experimentally validate NP:BGC pairings.



Metabologenomics datasets can inform direct cloning and heterologous expression for validating NP biosynthetic pathways. (A) 25 known NP:BGC relationships served as validation for metabologenomics-based GCF chemical annotations. NP:GCF pairs are represented by individual points. Known pairings are shown in purple, validating these associations. The pestalamides are shown in red. This analysis gives confidence in future metabologenomics efforts for establishing rich, mineable datasets for discovering and producing natural products of interest. (B) Cloning with CRISPR-based enzymes allows direct sequence-agnostic capture and manipulation of large BGCs, up to 150kb, and facilitates rapid heterologous expression with AMA1-containing vectors for elucidating and validating biosynthetic pathways. Here, the 38.4kb *pst* cluster was cloned from *Aspergillus brasiliensis* and heterologously expressed in *A. nidulans*. (C) Heterologous expression of the 37.9kb imizoquin (*imq*) cluster required overexpression of the in-cluster transcription factor to overcome innate transcriptional regulation.

	Overexpression Strategy	Genetic Stability	Titer Potential	Scalability/ Throughput	Technical Complexity	Cost/ Burden
New Approach	*Active Chromatin, Copy Number, & Strong Promoters*	*High*	*High*	*High*	*Low*	*Low*
Targeted Integration	Strong Promoters	High	Low	Low	High	High
Random Integration	Strong Promoters & Copy Number	Medium	High	High	Low	Medium
AMA1	Strong Promoters	Low	Low	High	Medium	Medium

A new fungal genome engineering approach for stable, high titer heterologous overexpression. Considering the strengths and weaknesses of modern fungal genome engineering techniques, optimal overexpression of heterologous products could most directly be accomplished through multiple integrations of optimized payloads into transcriptionally active chromatin. Terra Bioforge is working to refine such an approach to provide a platform, fed by metabologenomics and whole BGC cloning/manipulation expertise, to achieve higher natural product titers for functional characterization and overproduction.

References

- Robey MT, Caesar LK, Drott MT, Keller NP, Kelleher NL. An interpreted atlas of biosynthetic gene clusters from 1,000 fungal genomes. Proc Natl Acad Sci U S A. 2021 May 11;118(19):e2020230118. doi: 10.1073/pnas.2020230118. PMID: 33941694; PMCID: PMC8126772.
- Caesar LK, Butun FA, Robey MT, Ayon NJ, Gupta R, Dainko D, Bok JW, Nickles G, Stankey RJ, Johnson D, Mead DA, Cank KB, Earp CE, Raja HA, Oberlies NH, Keller NP, Kelleher NK. Correlative metabologenomics of 110 fungi reveals metabolite/gene cluster pairs. Nature Chemical Biology – *accepted for publication*
- Blin K, *et al*. antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Res. 2021 Jul 2;49(W1):W29-W35. doi: 10.1093/nar/gkab335. PMID: 33978755; PMCID: PMC8262755.
- van Santen JA, *et al*. The Natural Products Atlas: An Open Access Knowledge Base for Microbial Natural Products Discovery. ACS Cent Sci. 2019 Nov 27;5(11):1824-1833. doi: 10.1021/acscentsci.9b00806. Epub 2019 Nov 14. PMID: 31807684; PMCID: PMC6891855.