Unearth Novel Anti-MRSA Antibiotics from Extremophilic Fungi from a Toxic Mine Pit

Michael Swyers[1], Yansel Nunez[1], Md N. Islam[1], Jennifer Jaros[1], Rosa Ye[1], Andrea A. Stierle[2], Donald B. Stierle[2], Chengcang C. Wu[1]

[1]Intact Genomics Inc., St Louis, Missouri, USA. [2]Department of Biomedical and Pharmaceutical Sciences, University of Montana, Missoula, Montana, USA

Contact information: Intact Genomics, Inc., Phone: (314) 942-3655. Fax: (314) 942-3656. Email: cwu@intactgenomics.com. Website: www.intactgenomics.com

Abstract

The vast majority of fungal-encoded chemical space is uncharted due to limitations in culturing, *de-novo* sequencing, and/or difficulties in genetic manipulation of many fungi. We have developed a scalable platform using fungal artificial chromosomes (**FACs**) to capture full-length secondary metabolite biosynthetic gene clusters (SM-BGCs) derived from sequenced filamentous fungi[1] of *Aspergillus terreus, A. aculeatus, A. wentii, Fusarium solani, Penicillium expansum, Talaromyces marneffei* and *Pseudogymnoascus destructans*. Host *A. nidulans* strains transformed with FACs are screened by untargeted liquid chromatography-mass spectrometry (LC-MS) with ultrahigh mass accuracy**.** FAC-encoded products are recognized by a robust **FAC-MS** scoring system to identify spectral features most likely associated with each FAC[2]. By co-culturing *Penicillium fuscum* and *P. camembertii/clavigerum* isolated from the toxic Berkeley Pit, we have also discovered new macrolide compounds including berkeleylactone A, which holds a potentially novel mode of action for its antibiotic activity against 4 strains of Methicillin-resistant *Staphylococcus aureus* (MRSA), *Bacillus anthracis* (the anthrax bacterium), *Streptococcus pyogenes* (strep throat), *Candida albicans*, and *Candida glabrata* (pathogenic yeasts in humans)[3].

In this research, we will further integrate next-generation sequencing (NGS) with FAC-MS as a single **FAC-NGS-MS** platform to achieve 100kb-linked high-throughput sequencing and *de-novo* assembling of any fungal genome using only Illumina short reads. This platform also allows us to easily annotate fungal genomes and perform antiSMASH analysis. Moreover this technology has enabled capture of an entire set of full-length SM-BGCs of *P. fuscum* and *P. camembertii/clavigerum* for the rapid discovery of small molecule compounds including the above anti-MRSA antibiotics. We will discuss how this **FAC-NGS-MS** platform will enable scientists to easily move from identification of any unsequenced fungal novelty including from this unique toxic Berkeley pit to assigning metabolic and functional capabilities.

Unbiased shuttle BAC/FAC libraries from microbial genomes and metagenomes with average inserts >100 kb

Intact Genomics has constructed more than 200 unbiased Random Shear BAC libraries with large insert size from 100~180 kb for researchers around the world and arrayed a total of more than 2,500,000 clones from human, animals, plants, and other species. Our group has also constructed at least 56 unbiased Random Shear BAC libraries of bacteria, fungi, algae, soil and other environmental metagenome samples with average insert sizes of > 100kb (Figure 1, Table 1).

Unbiased shuttle BAC libraries

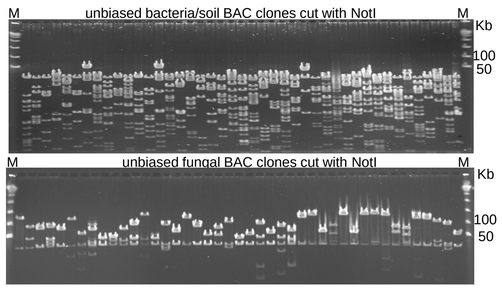


Figure 1. Genomic DNA was isolated from bacteria, fungi or environmental samples, randomly sheared, size-selected to >100 kb, and cloned into either shuttle vector: pIG-BAC4 or pIG-BAC5. DNA from minipreps was digested with NotI to excise inserts. The vector band is visible at about 15kb, and average insert size of this Random Shear shuttle BAC library of soil metagenomes was estimated at 120kb (top panel) and 125kb (bottom panel).

Table 1 Unbiased BAC libraries of microbes/metagenomes

Species (Samples)	No. of BAC/FAC libraries	Species/Sample names
Fungi	11	*Aspergillus terreus, A. aculeatus, A, wentii, Fusarium solani, F. virguliforme, Penicillium expansum, P. marneffei, Phytophthora infestans, Hyaloperonospora parasitica, Bremia lactucae, Geomyces destructans*
Algae	10	NA
Streptomyces strains	12	NA
Other bacterial strains	15	*Microadriaticum Strain CCMP 2467, E. coli O-157-H7, human pathogen, Citrobacter rodentium, E. coli STEC LB 226692, etc.*
Soil	10	NA
Other environmental samples	3	*Deep ocean metagenomics sample, 2 insect gut microbe metagenomes*
Total	**61**	**NA**

FAC enables ~100kb-linked sequencing with short-reads and better sequence assembly of un-sequenced extremophilic fungi

We have constructed unbiased shuttle FAC libraries (average 100kb or larger) from *Penicillium fuscum* (2A) and *P. camembertii/clavigerum* (2B) isolated from the toxic Berkeley Pit (not shown). In the first pass, we prepared the pooled FAC DNAs from each of ten 384-well plates per FAC library, an Illumina true-seq library for each FAC pool with an index, total 20 indexing Illumina true-seq libraries from the above 2 FAC libraries. We completed 2 MiSeq runs with v3 chemistry 2x300bp and generated ~32Gbp of sequencing data. By setting up an assembly, annotation, and antiSMASH pipeline for the sequencing data analysis, we were able to assemble the BAC pools as individual FACs (~1,000 contigs, >100kb each, ~2,500 contigs, >50kb each, Table 2) with only Illumina short-reads. Large-insert FACs provide 100kb-linked sequencing and make the assembly simple with high-quality results.

Table 2. Indexing FAC pool sequencing with short-reads only and assembly

FAC Pools	Raw Seq Data Generated (Mbp)	Genome Coverage (x)*	Total # Contigs	Avg. Contig Length (bp)
A1	1762.4	45.9	227	67,068
A2	1822.1	47.5	254	62,876
A3	1264.5	32.9	254	56,620
A4	1788.9	46.6	234	67,799
A5	1379.0	35.9	252	54,638
A6	1515.1	39.5	261	59,404
A7	1964.8	51.2	230	66,916
A8	1810.5	47.1	235	67,830
A9	1815.5	47.3	253	63,707
A10	1321.0	34.4	521	29,740
B1	1389.0	36.2	259	94,665
B2	1720.9	44.8	374	53,191
B3	1765.7	46.0	263	71,988
B4	1642.9	42.8	248	76,997
B5	1848.7	48.1	256	71,810
B6	1666.1	43.4	405	43,257
B7	1657.8	43.2	265	69,247
B8	1609.7	41.9	264	68,123
B9	1710.9	44.6	237	79,225
B10	1403.9	36.6	314	60,649

*Genome coverage was calculated based on 384 FAC clones per plate, and average 100kb per FAC clone.

FAC bridges the gap between sequencing and functional study of un-sequenced extremophilic fungi

Through annotation and antiSMASH analysis, we discovered at least 19 full-length BGCs from 2A and 25 full-length BGCS from 2B in the first pass (Table 3). In this process, we not only acquired the sequencing and annotation data, including predicted BGCs, but also obtained the sequenced FACs and full-length BGC-FACs data. These are now ready for the heterologous host (*A. nidulans)* transformation, expression, and compound identification. Therefore, FAC-NGS is a powerful semisynthetic tool.

A. nidulans transformation with large 100kb BGC-FACs is very challenging. Initially, we developed a working protocol to prepare and purify high-quality *A. nidulans* spheroplasts for transformation, but we were able to formulate and optimize an enzyme-treatment protocol without the need of purification of the spheroplast for FAC transformation. We can routinely obtain ~500 transformants per transformation by this simplified *A. nidulans* BGC-FAC transformation method (Figure 2).

Table 3. Predicted SM gene clusters in 2 fungi.

SM Clusters predicted	*P. fuscum*	*P. camembertii/ clavigerum*
DMAT	0	2
HYBRID	2	5
NRPS	4	4
PKS	21	22
Total	**≥19**	**≥25**

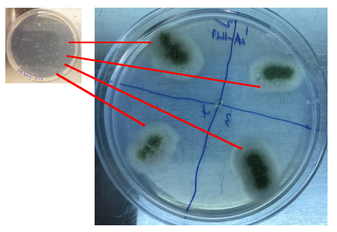


Figure 2. Re-inoculating individual *A. nidulans* (An) FAC transformants growing on selective GMM media without uridine and uracil. The heterologous host *A. nidulans* strain: RJW256.2 is pyrO$^-$ and pyrG$^-$ mutant and does not grow on the selective GMM media plate. The FAC vector provides pyrG$^+$ gene. Therefore, FACs transformants can grow on the selective GMM media plate without uridine and uracil.

Anti-MRSA compound-berkeleylactone A from extremophilic fungi

By co-culturing *Penicillium fuscum* and *P. camembertii/clavigerum* isolated from the toxic Berkeley Pit, we also discovered new macrolide compounds including berkeleylactone A (Figure 3) which holds a potentially novel mode of action for its antibiotic activity against 4 strains of Methicillin-resistant *Staphylococcus aureus* (MRSA), *Bacillus anthracis* (the anthrax bacterium), *Streptococcus pyogenes* (strep throat), *Candida albicans*, and *Candida glabrata* (pathogenic yeasts in humans)[3]

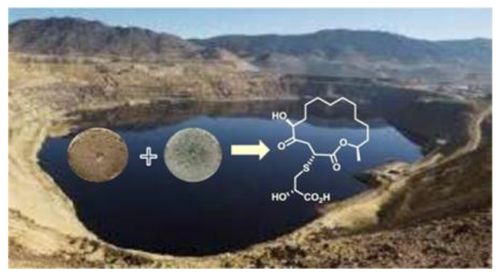
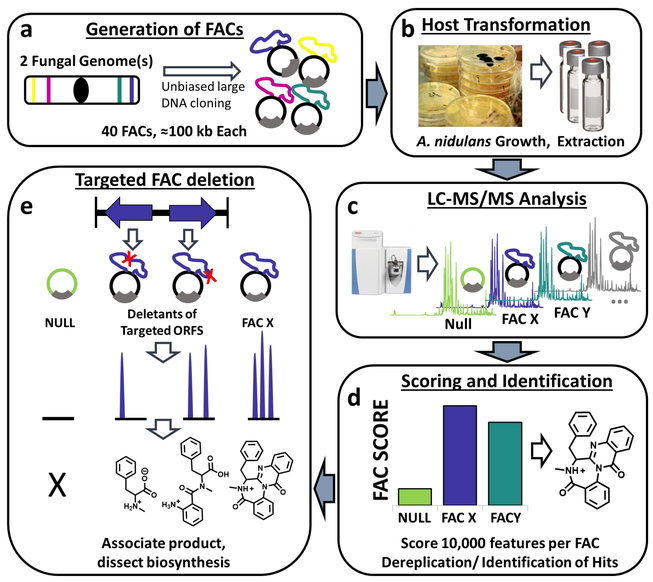


Figure 3. Illustration of new macrolide compounds including berkeleylactone A discovered by co-culturing 2 extremophilic filamentous fungi from the toxic Berkeley Pit.

FAC-NGS-MS Pipeline



We have successfully invented the FAC-NGS-MS platform: a novel scalable technology with potential to revive the fungal natural product discovery and deliver solutions for wide-range bioenergy, environmental microbiology, agriculture, pharmaceutical and other industries (Figure 4).

Figure 4. The **FAC-MS** pipeline. A) Generation of FACS from randomly sheared genomic DNA and cloning into self-replicating *E. coli*- fungal shuttle vectors-NGS sequencing. B) Transformation into expression host, *A. nidulans*, and extraction of secondary metabolites (SMs). C) Untargeted high resolution LC-MS/MS analysis of SM extracts from each FAC. D) Feature detection, scoring, and hit identification/ dereplication. E) Generation of FAC deletants, association of SMs with FAC ORFs, dissection of biosynthetic pathway.

Conclusion

Intact Genomics is the sole worldwide provider of high-quality, unbiased BAC/FAC libraries. We have constructed random shear shuttle BAC/FAC libraries from bacteria, fungi and metagenomic samples with an average insert size of at least 100kb, large enough to contain entire sets of full-length PKS/NRPS pathways. This technology enables natural product discovery on a scale not possible before. We also provide custom services and are open to collaborations on the development of functional genomics, synthetic biology, and natural product discovery research.

References

[1] Bok JW, Ye R, Clevenger KD, Mead D, Wagner M, Krerowicz A, Albright JC, Goering AW, Thomas PM, Kelleher NL, Keller NP, Wu CC. Fungal artificial chromosomes for mining of the fungal secondary metabolome. BMC Genomics. 2015 Apr 29;16(1):343. [2] Clevenger KD, Bok JW, Ye R, Miley GP, Verdan MH, Velk T, Chen C, Yang K, Robey MT, Gao P, Lamprecht M, Thomas PM, Islam MN, Palmer JM, Wu CC, Keller NP, Kelleher NL. A scalable platform to identify fungal secondary metabolites and their gene clusters. Nat Chem Biol. 2017 Aug;13(8):895-901. [3] Stierle AA, Stierle DB, Decato D, Priestley ND, Alverson JB, Hoody J, McGrath K, Klepacki D. The Berkeleylactones, Antibiotic Macrolides from Fungal Coculture. J Nat Prod. 2017 Apr 28;80(4):1150-1160.

Acknowledgments

We thank Drs. Igor Grigoriev, Nancy Keller, Neil Kelleher for their generous support and many collaborators worldwide from our custom service projects. This research is partially supported by the NIH SBIR programs: R43AI138765, R44AT009158.

Intact Genomics®

UNIVERSITY OF MONTANA

National Institute of Allergy and Infectious Diseases

NIH National Center for Complementary and Integrative Health