Lawrence Berkeley National Laboratory

Recent Work

Title

IMG-ABC: An Atlas of Biosynthetic Gene Clusters to Fuel the Discovery of Novel Secondary Metabolites

Permalink

https://escholarship.org/uc/item/9b64p7jh

Authors

Chen, I-Min Chu, Ken Ratner, Anna et al.

Publication Date

2014-10-28

IMG-ABC: An <u>A</u>tlas of <u>B</u>iosynthetic Gene <u>C</u>lusters to Fuel the Discovery of Novel Secondary Metabolites

Authors: *Michalis Hadjithomas¹, I-Min Chen², Ken Chu², Anna Ratner², Krishna Palaniappan², Jinghua Huang², T.B.K. Reddy¹, Peter Cimermancic³, Michael A. Fishchbach³, Natalia Ivanova¹, Victor Markowitz¹, Nikos Kyrpides¹, *Amrita Pati¹

¹ U.S. Department of Energy Joint Genome Institute // LBNL - Walnut Creek, CA
² Biological management and Technology Center, Computational Research Division // LBNL – Berkeley, CA

October, 2014

ACKNOWLEDGMENTS:

Work by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. This work was supported by the Office of Biological and Environmental Research, Life Sciences Division, U.S. Department of Energy (Contract No. [DE-AC02-05CH11231]). This research used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy (Contract No. [DE-AC02-05CH11231]). We thank Dr. Thomas Girke, Tyler Backman and Kevin Horan at UC Riverside for their assistance and the implementation of the ChemmineR package and Dr. Marnix Medema for his support of AntiSMASH

DISCLAIMER:

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial

³ Department of Bioengineering and Therapeutic Sciences // University of California-San Francisco – San Francisco, CA

^{*} To whom correspondence may be addressed. Amrita Pati, Ph.D., Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA, 94598, USA - <u>APati@lbl.gov</u> or Michalis Hadjithomas, Ph.D., Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA, 94598, USA – <u>Michalis@lbl.gov</u>

IMG-ABC: An <u>A</u>tlas of <u>B</u>iosynthetic Gene <u>C</u>lusters to Fuel the Discovery of Novel Secondary Metabolites

product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.



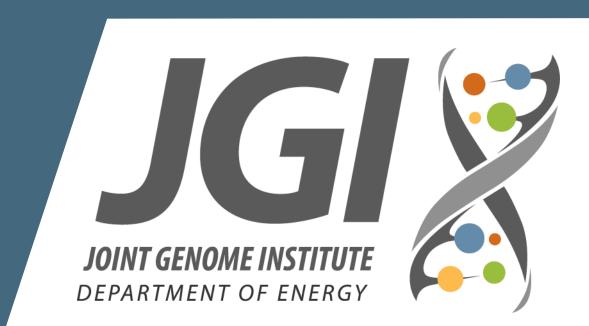
IMG-ABC: An Atlas of Biosynthetic Gene Clusters to Fuel the Discovery of Novel Secondary Metabolites

Michalis Hadjithomas¹, I-Min Amy Chen², Ken Chu², Anna Ratner², Krishna Palaniappan², Jinghua Huang², T. B. K. Reddy¹,

Peter Cimermančič³, Michael A. Fischbach³, Natalia N. Ivanova¹, Victor Markowitz², Nikos C. Kyrpides¹, Amrita Pati¹

¹ Prokaryotic Super Program, DOE Joint Genome Institute, Walnut Creek, CA, ² Biological Data Management and Technology Center, Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA,

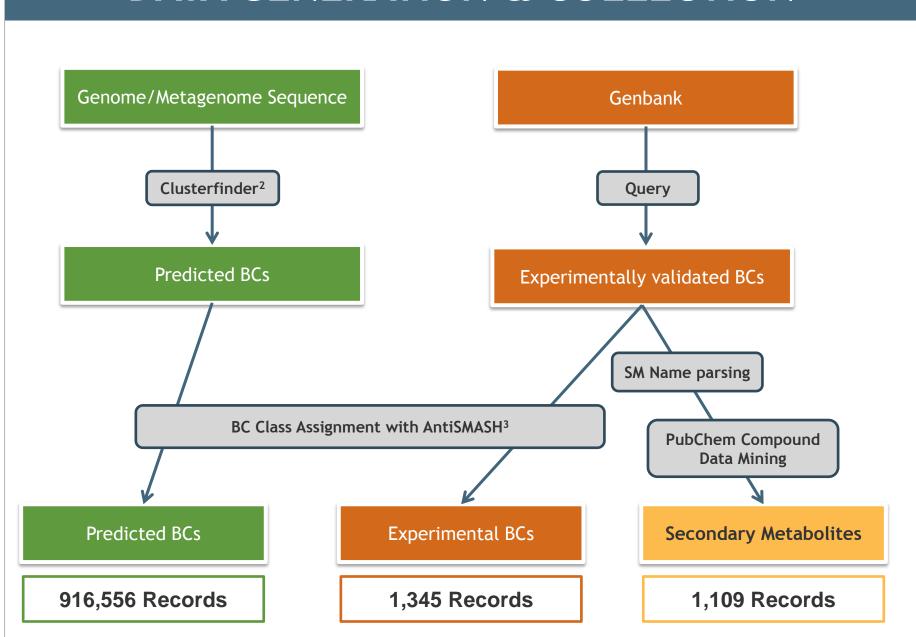
³ Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco CA



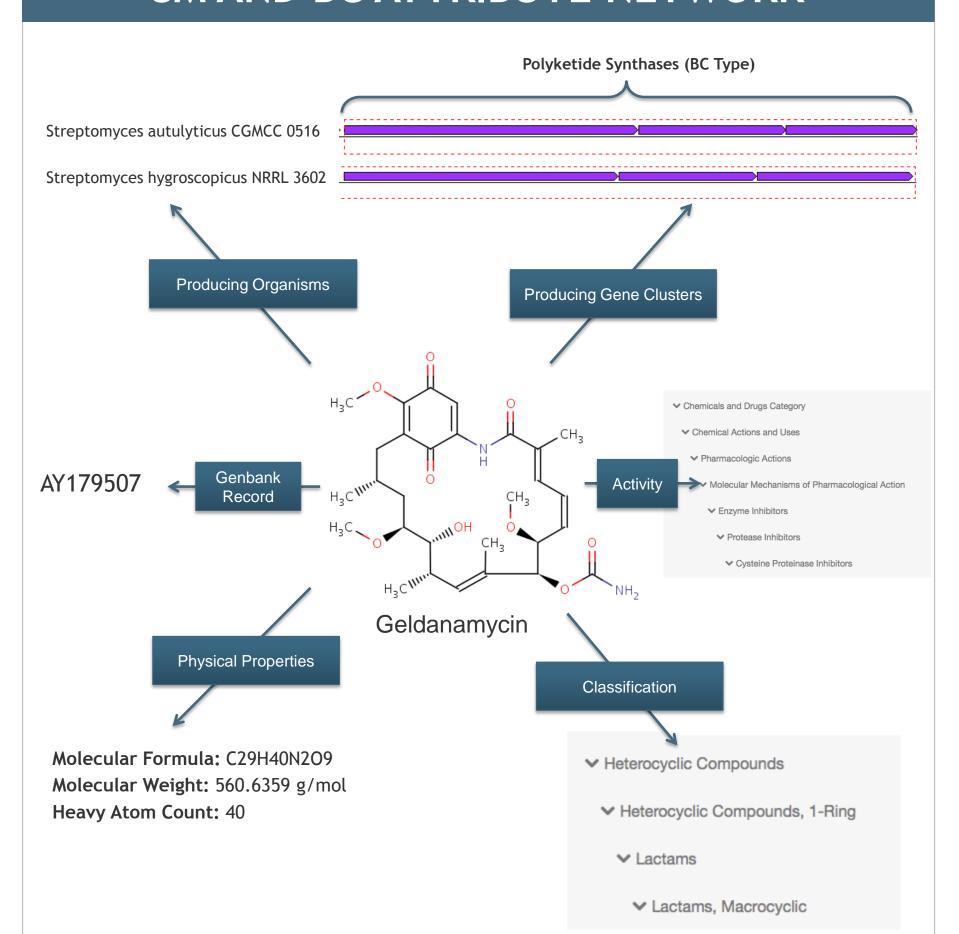
INTRODUCTION

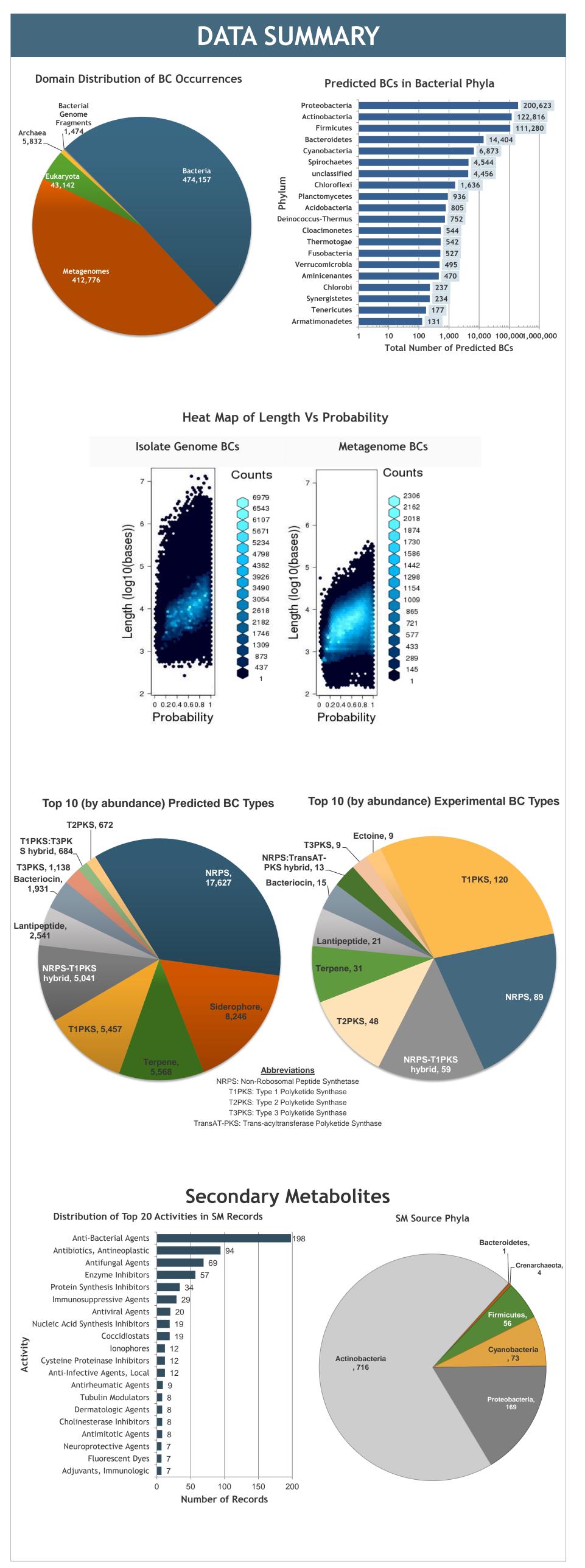
In the discovery of secondary metabolites (SMs), large-scale analysis of sequence data is a promising exploration path that remains largely underutilized due to the lack of relevant computational resources. We present IMG-ABC (https://img.jgi.doe.gov/abc/) -- An Altlas of Biosynthetic gene Clusters within the Integrated Microbial Genomes (IMG) system1. IMG-ABC is a rich repository of both validated and predicted biosynthetic clusters (BCs) in cultured isolates, single-cells and metagenomes linked with the SM chemicals they produce and enhanced with focused analysis tools within IMG. The underlying scalable framework enables traversal of phylogenetic dark matter and chemical structure space -- serving as a doorway to a new era in the discovery of novel molecules.

DATA GENERATION & COLLECTION



SM AND BC ATTRIBUTE NETWORK



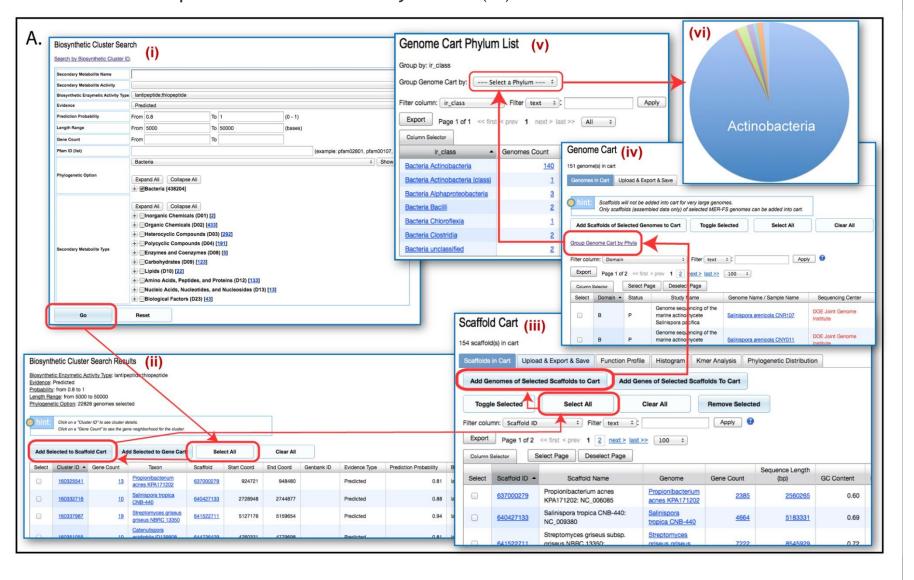


USER INTERFACE AND ANALYSIS TOOLS IMG-ABC can be accessed directly at https://img.jgi.doe.gov/abc/ or through the ABC tab in the IMG system. Biosynthetic Gene Clusters Clusters of genes whose expression leads Search BC/SM by ID **BC** Biosynthetic Clusters Search BCs CONCERNATION OF THE PROPERTY O Search SMs Continues Sentential Sentential Sentential per Color, Space SCI (1981) ACCORD Company of the Control of the Contro Browse SMs **BC** Detail Page Genome: Stigmatella aurantiaca DW4/3-1 Cluster ID: 160309465 Genes in Cluster | Cluster Neighborhood | Secondary Metabolite | IMG Pathways | Metacyc | KEGG **Biosynthetic Cluster Information** Stigmatella aurantiaca DW4/3-1 Project ID: Gc01414 Scaffold Stigmatella aurantiaca DW4/3-1 chromosome: NC_01462 **DNA Sequence** 5893433..5944441 START_ON_CHROMOSOME BC_TYPE bacteriocin;nrps;t1pks END_ON_CHROMOSOME EVIDENCE PROBABILITY GENBANK ACC PFAM_COUNT GENE_COUNT LENGTH **SM Detail Page** Object ID es peucetius (Biosynthetic Cluster 160962528) eucetius ATCC 29050 (Biosynthetic Cluster 1609623 Common Name s peucetius ATCC 29050 (Biosynthetic Cluster 16096261) Ext Accession | C01661 myces peucetius (Biosynthetic Cluster 16096235) ces peucetius (Biosynthetic Cluster 160962521) us ATCC 29050 (Biosynthetic Cluster 16096238 omyces peucetius (Biosynthetic Cluster 160962613) C27H29NO11 Image from NCI/CADD. Source url. CAS: 23214-92-8 InChl=1S/C27H29NO11/c1-10-22(31)13(28)6-17(38-10)39-15-8-27(36,16(30)9-29)7-12-19(15)26(35)2 /h3-5,10,13,15,17,22,29,31,33,35-36H,6-9,28H2,1-2H3/t10-,13-,15-,17-,22+,27-/m0/s AOJJSUZBOXZQNB-TZSSRYMLSA-N Search by BC and SM Attributes IMG/M ER Home Find Genomes Find Genes Find Functions Compare Genomes Analysis Cart OMICS ABC My IMG Biosynthetic Cluster (BC) Search Search Biosynthetic Cluster by ID Secondary Metabolite Name Secondary Metabolite Activity Length Range Gene Count Pfam ID (list) (example: pfam02801, pfam00107, etc Inorganic Chemicals (D01) [SM:1 BC:1] + Organic Chemicals (D02) [SM:410 BC:345] + Heterocyclic Compounds (D03) [SM:280 BC:24 Search by SM Structure + Polycyclic Compounds (D04) [SM:282 BC:311] Hormones, Hormone Substitutes, and Hormor + Enzymes and Coenzymes (D08) [SM:6 BC:6] Search By Chemical Structure + Carbohydrates (D09) [SM:147 BC:128] + Lipids (D10) [SM:33 BC:36] + Amino Acids, Peptides, and Proteins (D12) [SI + Nucleic Acids, Nucleotides, and Nucleosides + Biological Factors (D23) [SM:46 BC:78] Similarity Cutoff: 0.5 (value between 0 and 1, rounded to 2 decimal places)

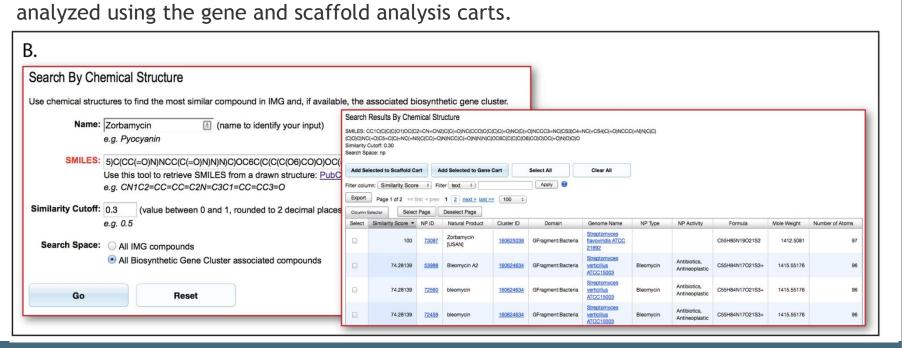
Go Reset

ANALYSIS WORKFLOW EXAMPLE

(A) Using the search functions of IMG-ABC in combination with IMG's analysis tools to investigate the distribution of Biosynthetic Gene Clusters (BCs) of specific classes. (i) A search for BC with desired attributes is performed through the BC search interface results and yields a (ii), list of records, which can be added (iii) first to IMG's Scaffold analysis cart and then (iv) to the Genome analysis cart. The organisms harboring BCs matching the initial search can be grouped (v) into different phylogenetic classes. These results can be exported for further analysis and (vi) visualization.



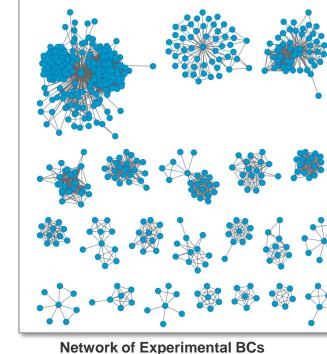
(B) An example of a search using a chemical structure as a query. The chemical structure descriptor string for Zorbamycin is used to search IMG-ABC for secondary metabolites (SM) that match the query with a similarity score greater than 0.3. This yields a list of 118 compounds sorted by descending similarity score whose associated BCs can be



PLANNED DEVELOPMENTS

IMG-ABC is continuously maintained and improved. Planned developments include:

- Applying algorithms to predict the SM backbone of certain classes of predicted BCs, and include these structures in a searchable database.
- Incorporating module annotation for specific classes of biosynthetic enzymes.
- Implementing a BC search by using a global BC pairwise similarity algorithm developed by the JGI – OMICS group.
- Using similarity algorithm to build an interactive BC network to facilitate novel BC/SM discovery



REFERENCES

- 1. Markowitz VM, Chen I-MA, Chu K, Szeto E, Palaniappan K, Pillay M, Ratner A, Huang J, Pagani I, Tringe S, et al. 2014. IMG/M 4 version of the integrated metagenome comparative analysis system. Nucleic Acids Res. 42:D568-D573.
- 2. Cimermancic P, Medema MH, Claesen J, Kurita K, Wieland Brown LC, Mavrommatis K, Pati A, Godfrey PA, Koehrsen M, Clardy J, et al. 2014. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. Cell 158:412-421.
- 3. Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, Takano E, Weber T. 2013. antiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers. Nucleic Acids Res. 41:W204-W212.

ACKNOWLEDGEMENTS

This work was supported by the Office of Biological and Environmental Research, Life Sciences Division, U.S. Department of Energy (Contract No. [DE-AC02-05CH11231]). This research used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy (Contract No. [DE-AC02-05CH11231]). We thank Dr. Thomas Girke, Tyler Backman and Kevin Horan at UC Riverside for their assistance and the implementation of the ChemmineR package and Dr. Marnix Medema for his support of AntiSMASH.

CONTACT

Amrita Pati, Ph.D. <u>apati@lbl.gov</u>
Michalis Hadjithomas, Ph.D. <u>michalis@lbl.gov</u>