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

Authors: *Michalis Hadjithomas[1], I-Min Chen[2], Ken Chu[2], Anna Ratner[2], Krishna Palaniappan[2], Jinghua Huang[2], T.B.K. Reddy[1], Peter Cimermancic[3], Michael A. Fishchbach[3], Natalia Ivanova[1], Victor Markowitz[1], Nikos Kyrpides[1], *Amrita Pati[1]

[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

[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 - APati@lbl.gov or Michalis Hadjithomas, Ph.D., Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA, 94598, USA – Michalis@lbl.gov*

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

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

Michalis Hadjithomas[1], I-Min Amy Chen[2], Ken Chu[2], Anna Ratner[2], Krishna Palaniappan[2], Jinghua Huang[2], T. B. K. Reddy[1], Peter Cimermančič[3], Michael A. Fischbach[3], Natalia N. Ivanova[1], Victor Markowitz[2], Nikos C. Kyrpides[1], Amrita Pati[1]

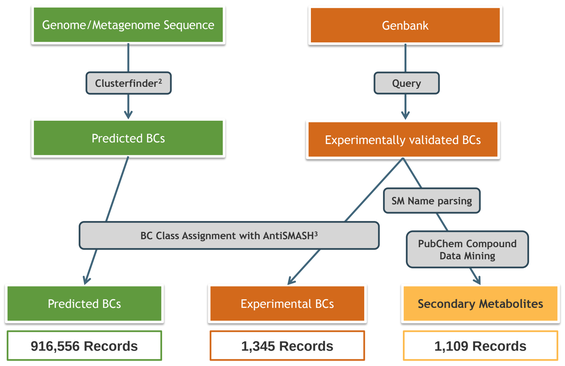
[1] Prokaryotic Super Program, DOE Joint Genome Institute, Walnut Creek, CA, [2] Biological Data Management and Technology Center, Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA,
[3] Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco CA

JOINT GENOME INSTITUTE
DEPARTMENT OF ENERGY

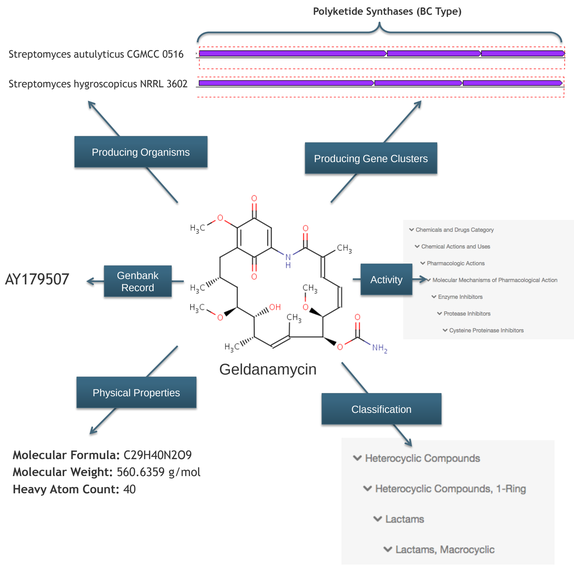
INTRODUCTION

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 Atlas of Biosynthetic gene Clusters within the Integrated Microbial Genomes (IMG) system[1]. 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.

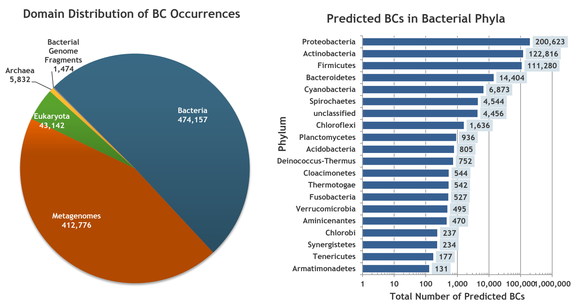
DATA GENERATION & COLLECTION

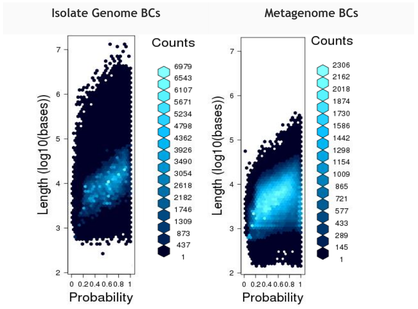


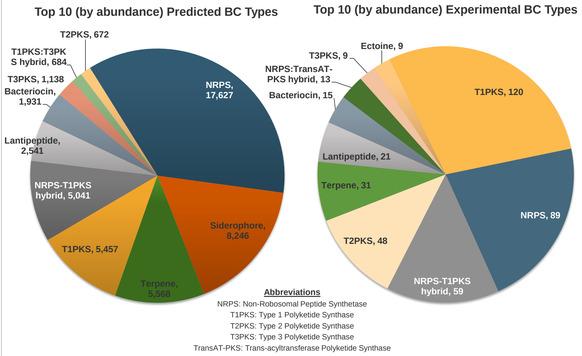
SM AND BC ATTRIBUTE NETWORK



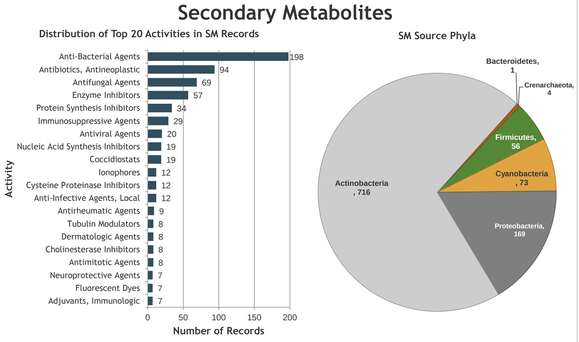
DATA SUMMARY





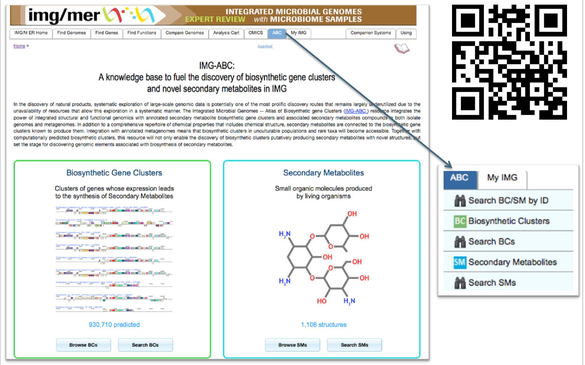


Abbreviations
NRPS: Non-Robosomal Peptide Synthetase
T1PKS: Type 1 Polyketide Synthase
T2PKS: Type 2 Polyketide Synthase
T3PKS: Type 3 Polyketide Synthase
TransAT-PKS: Trans-acyltransferase Polyketide Synthase



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.



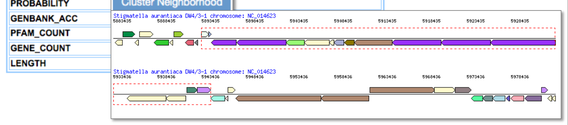
Genome: Stigmatella aurantiaca DW4/3-1
Cluster ID: 160309465

BC Detail Page

Biosynthetic Cluster | Genes in Cluster | Cluster Neighborhood | Secondary Metabolite | IMG Pathways | Metacyc | KEGG

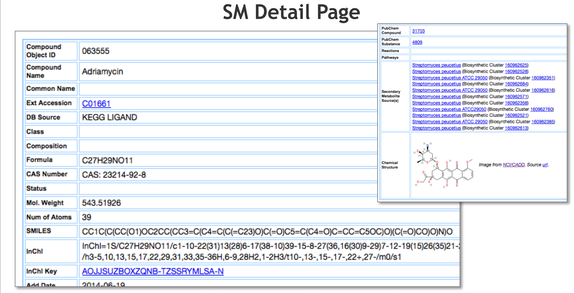
Biosynthetic Cluster Information	
Cluster ID	160309465
Genome	Stigmatella aurantiaca DW4/3-1
GOLD ID	Project ID: Gc01414
Scaffold	Stigmatella aurantiaca DW4/3-1 chromosome: NC_014623
DNA Sequence	5893433..5944441
START_ON_CHROMOSOME	5893433
BC_TYPE	bacteriocin;nrps;t1pks
END_ON_CHROMOSOME	5944441
EVIDENCE	
PROBABILITY	
GENBANK_ACC	
PFAM_COUNT	
GENE_COUNT	
LENGTH	

Cluster Neighborhood

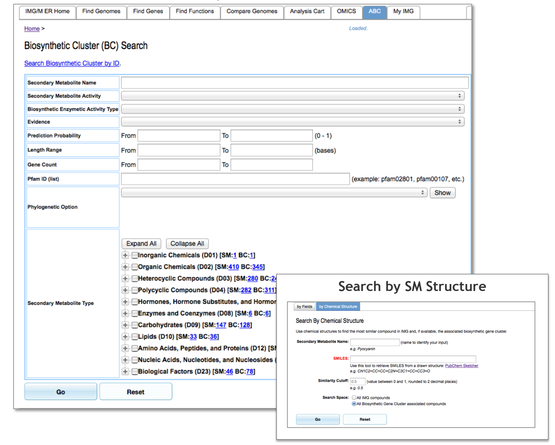


SM Detail Page

Compound Object ID	063555
Compound Name	Adriamycin
Common Name	
Ext Accession	C01661
DB Source	KEGG LIGAND
Class	
Composition	
Formula	C27H29NO11
CAS Number	CAS: 23214-92-8
Status	
Mol. Weight	543.51926
Num of Atoms	39
SMILES	CC1C(C(CC(O1)OC2CC(CC3=C(C4=C(C(=C23)O)C(=O)C5=C(C4=O)C=CC=C5OC)O)(C(=O)CO)O)N)O
InChI	InChI=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)21-24(33)14-5-3-4-11(37-2)18(14)25(34)20(21)23(12)32/h3-5,10,13,15,17,22,29,31,33,35-36H,6-9,28H2,1-2H3/t10-,13-,15-,17-,22+,27-/m0/s1
InChI Key	AOJJSUZBOXZQNB-TZSSRYMLSA-N
Add Date	2014-06-19



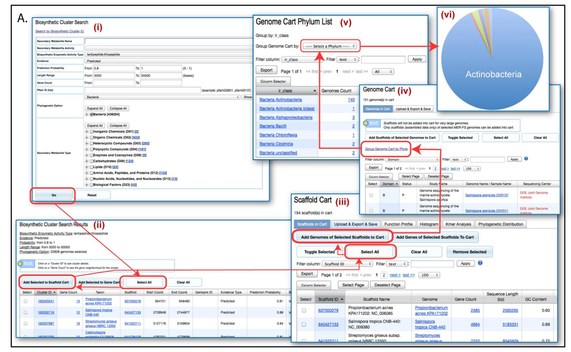
Search by BC and SM Attributes



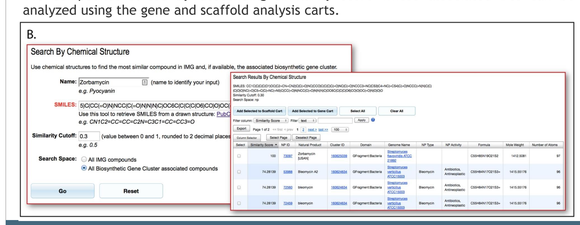
Search by SM Structure

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.



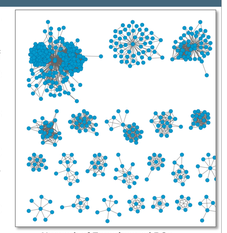
(B) An example of a search using a chemical structure as a query. The chemical structure descriptor string for Zorbamycin is used to search IMG-ABC for secondary metabolites (SM) that match the query with a similarity score greater than 0.3. This yields a list of 118 compounds sorted by descending similarity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



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



Network of Experimental BCs

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. apati@lbl.gov
Michalis Hadjithomas, Ph.D. michalis@lbl.gov

The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported under Contract No. DE-AC02-05CH11231.