

# 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 Atlas of Biosynthetic Gene Clusters to Fuel the Discovery of Novel Secondary Metabolites

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

<sup>1</sup> U.S. Department of Energy Joint Genome Institute // LBNL - Walnut Creek, CA

<sup>2</sup> Biological management and Technology Center, Computational Research Division // LBNL – Berkeley, CA

<sup>3</sup> 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](mailto:APati@lbl.gov) or Michalis Hadjithomas, Ph.D., Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA, 94598, USA – [Michalis@lbl.gov](mailto: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

## 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

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

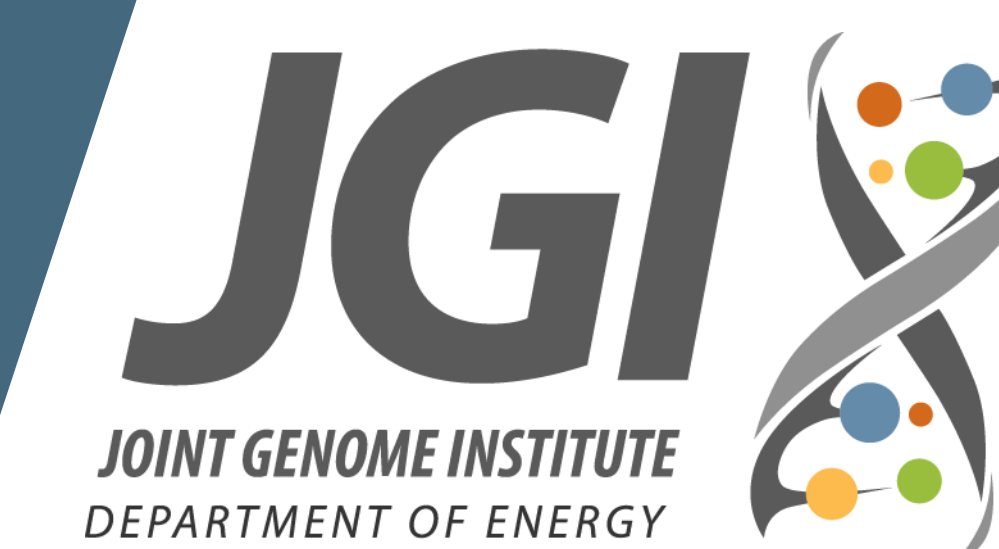
product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of 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 Hadjithomas<sup>1</sup>, I-Min Amy Chen<sup>2</sup>, Ken Chu<sup>2</sup>, Anna Ratner<sup>2</sup>, Krishna Palaniappan<sup>2</sup>, Jinghua Huang<sup>2</sup>, T. B. K. Reddy<sup>1</sup>, Peter Cimermančić<sup>3</sup>, Michael A. Fischbach<sup>3</sup>, Natalia N. Ivanova<sup>1</sup>, Victor Markowitz<sup>2</sup>, Nikos K. Kyriades<sup>1</sup>, Amrita Pati<sup>1</sup>

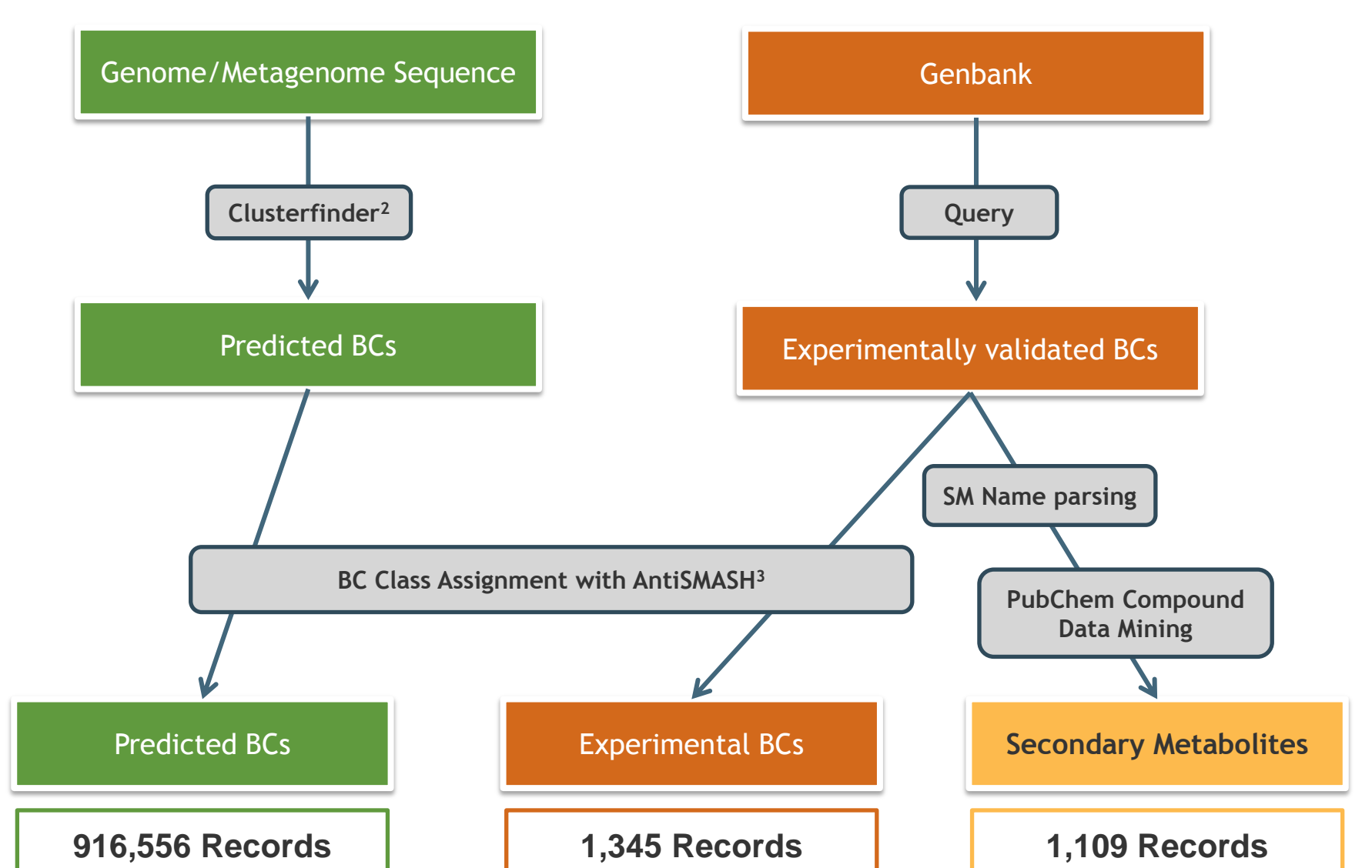
<sup>1</sup> Prokaryotic Super Program, DOE Joint Genome Institute, Walnut Creek, CA, <sup>2</sup> Biological Data Management and Technology Center, Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA, <sup>3</sup> Department of Bioengineering and Therapeutic Sciences,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



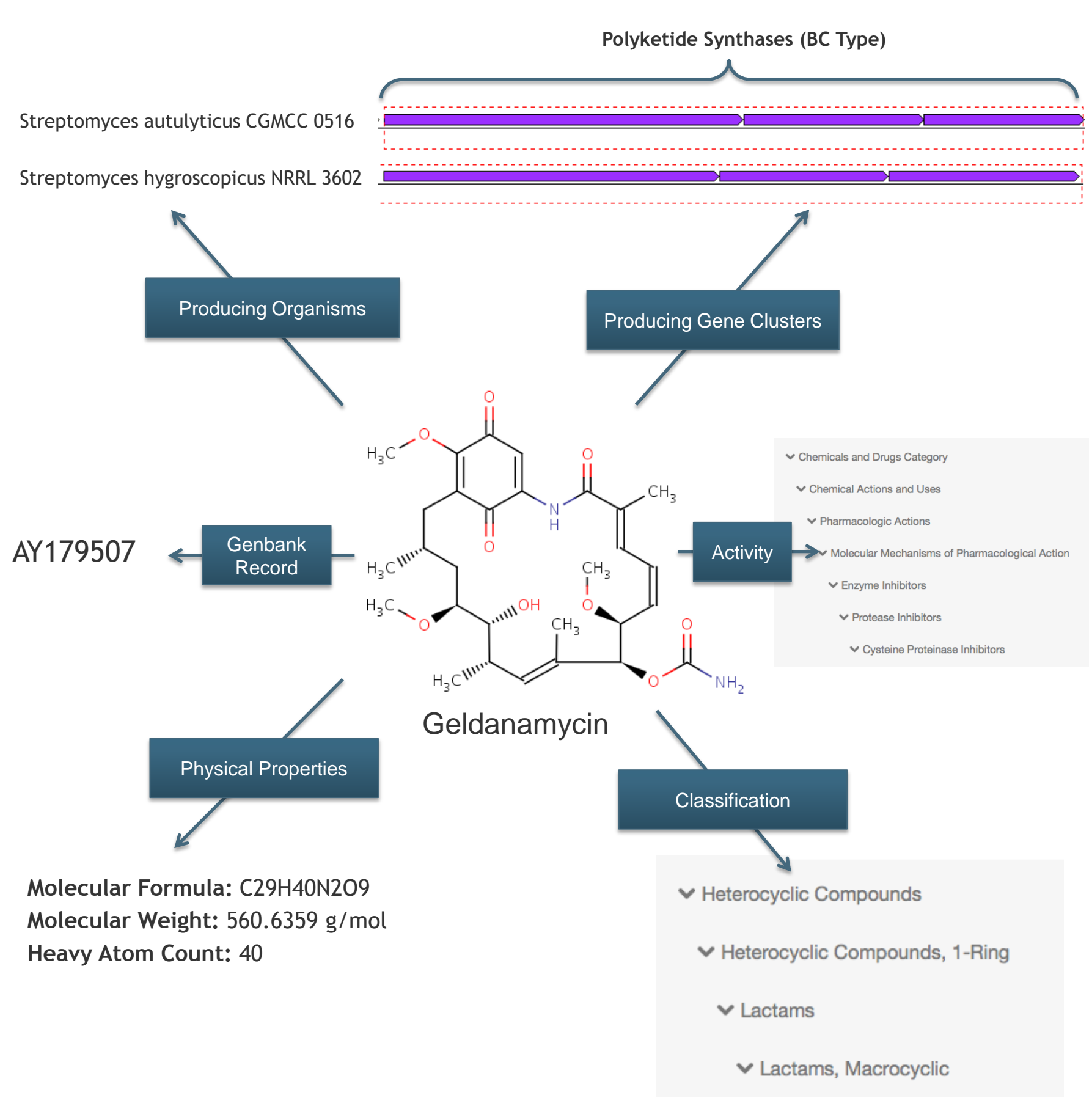
## INTRODUCTION

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 Atlas of Biosynthetic Gene Clusters within the Integrated Microbial Genomes (IMG) system<sup>1</sup>. 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.

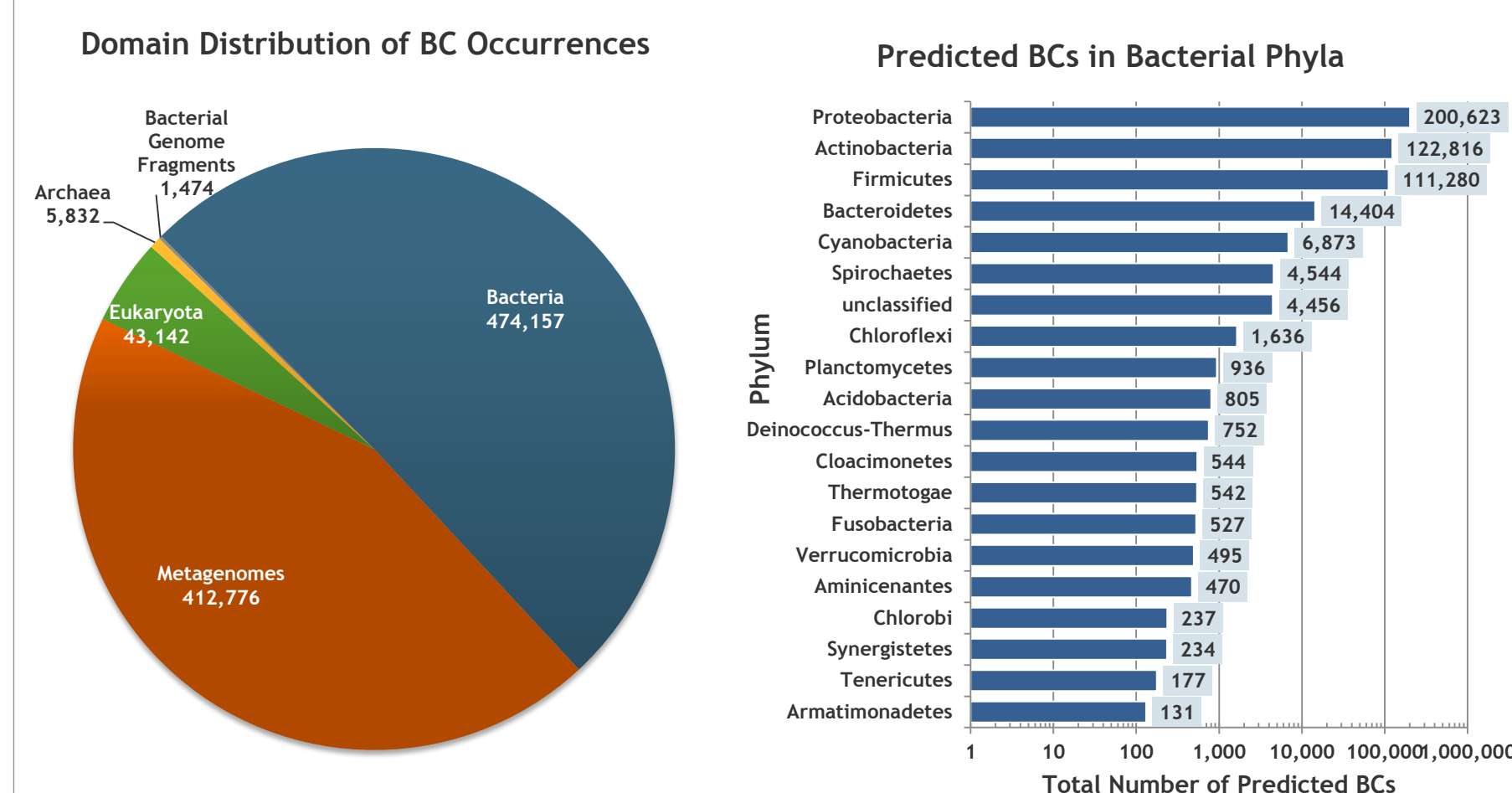
## DATA GENERATION & COLLECTION



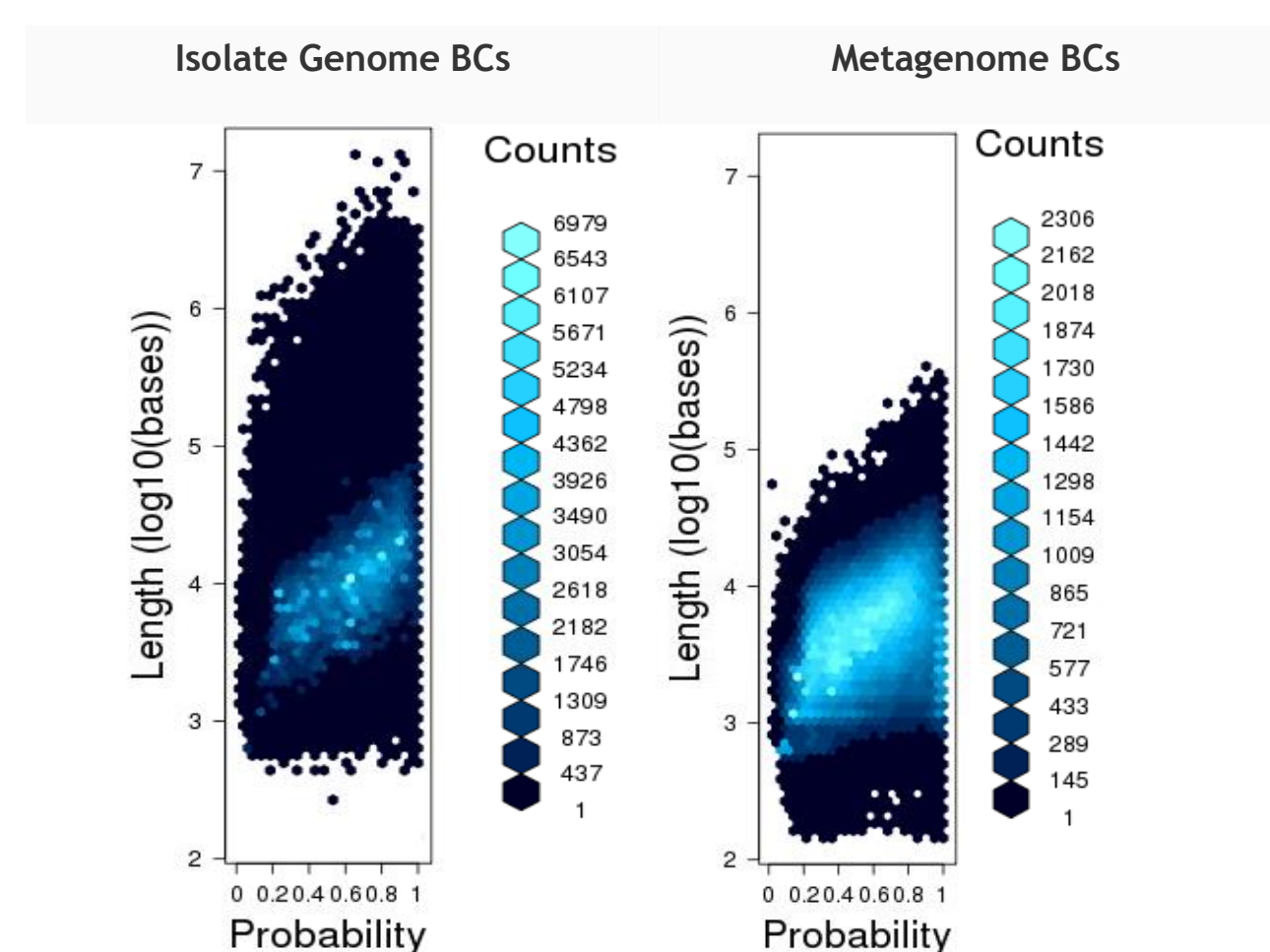
## SM AND BC ATTRIBUTE NETWORK



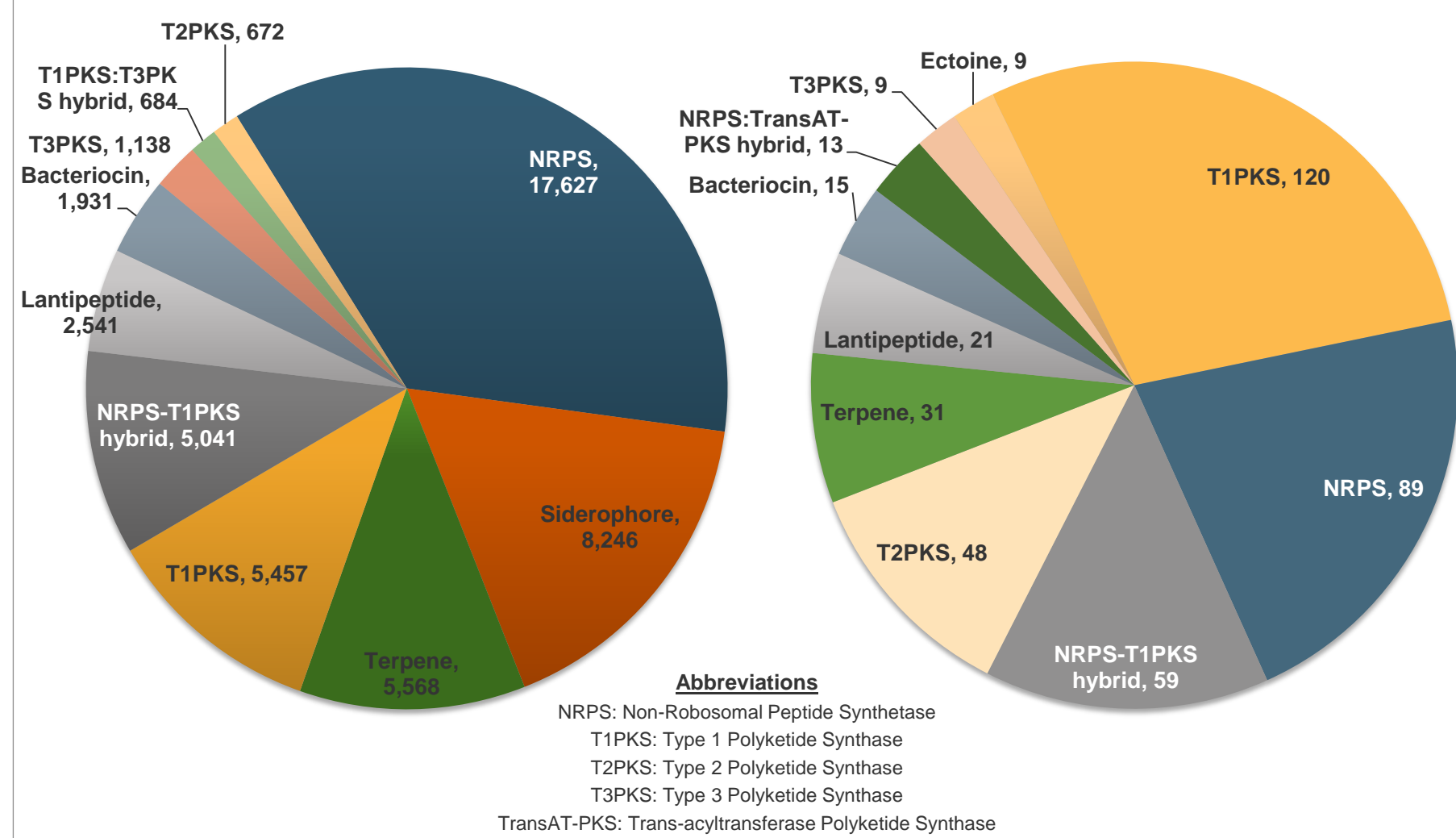
## DATA SUMMARY



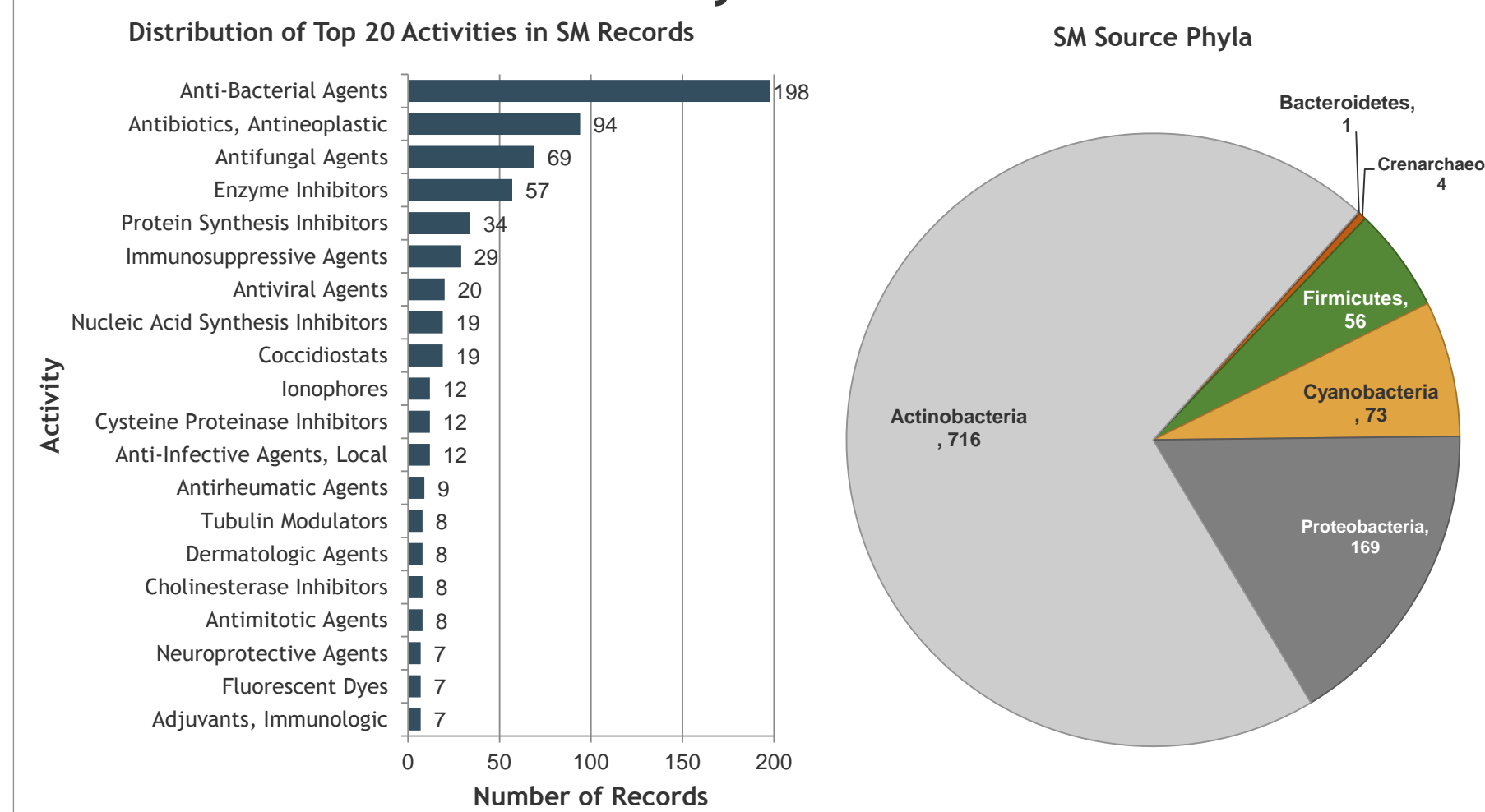
## Heat Map of Length Vs Probability



## Top 10 (by abundance) Predicted BC Types and Top 10 (by abundance) Experimental BC Types

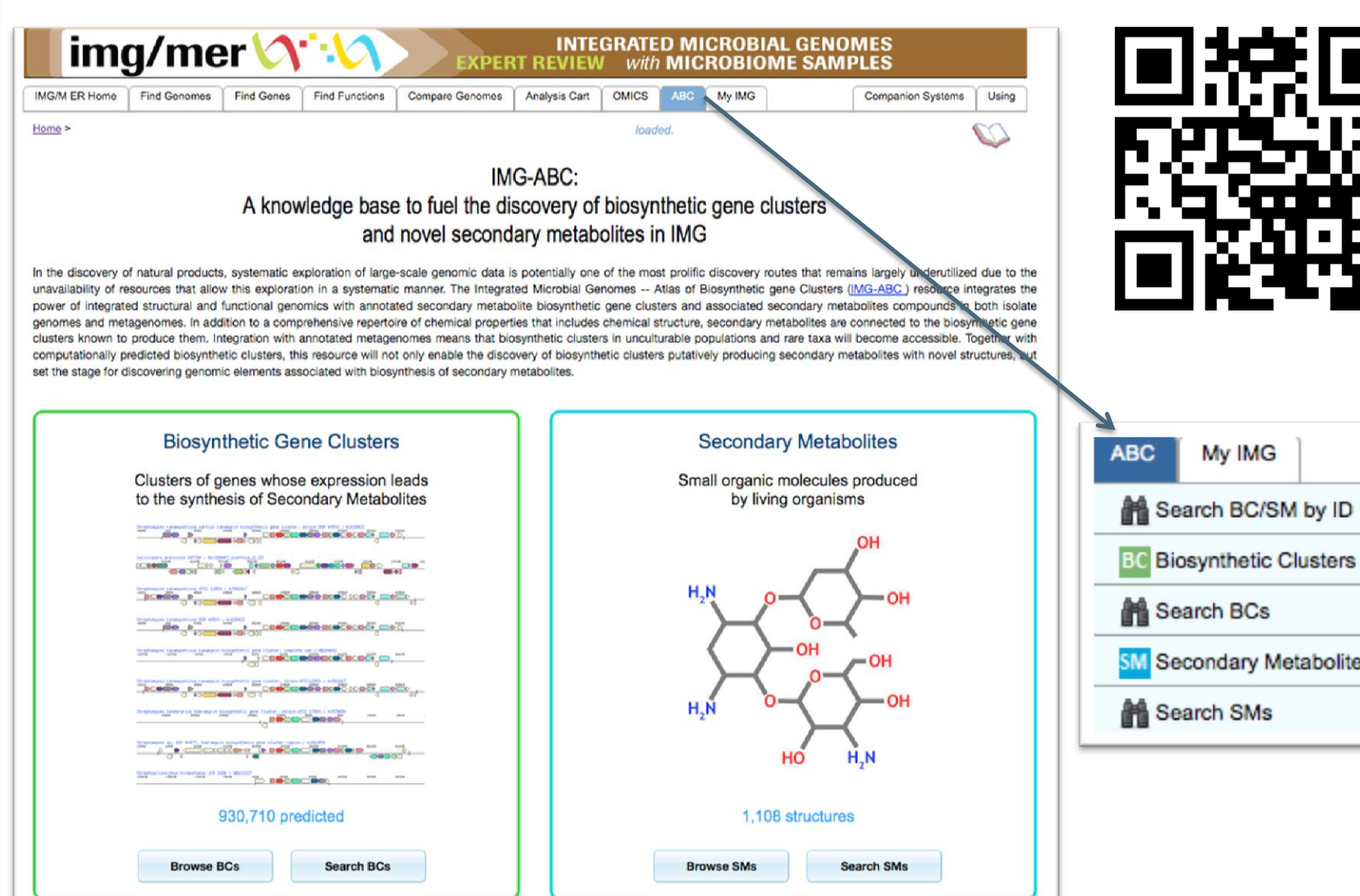


## Secondary Metabolites

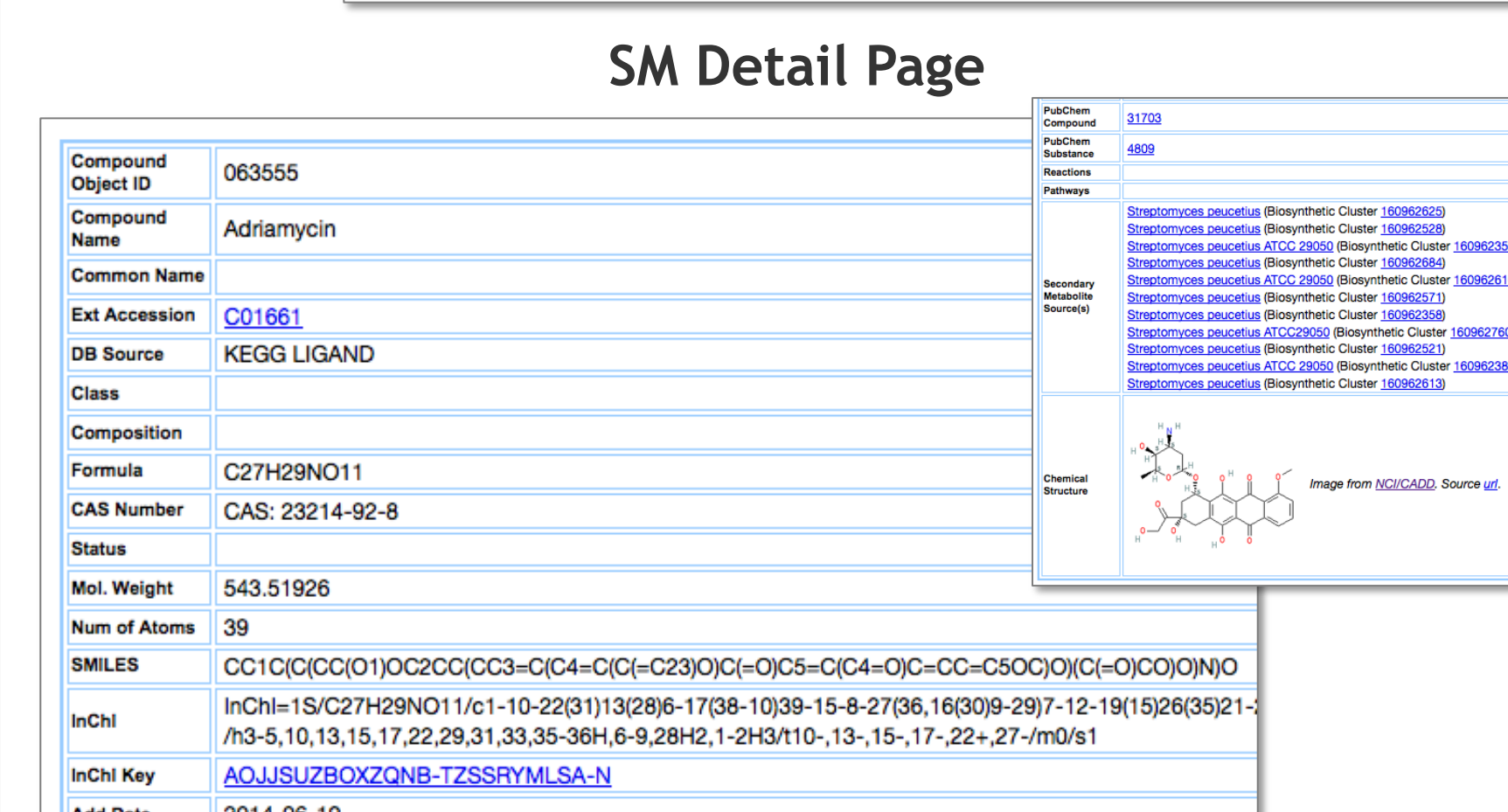
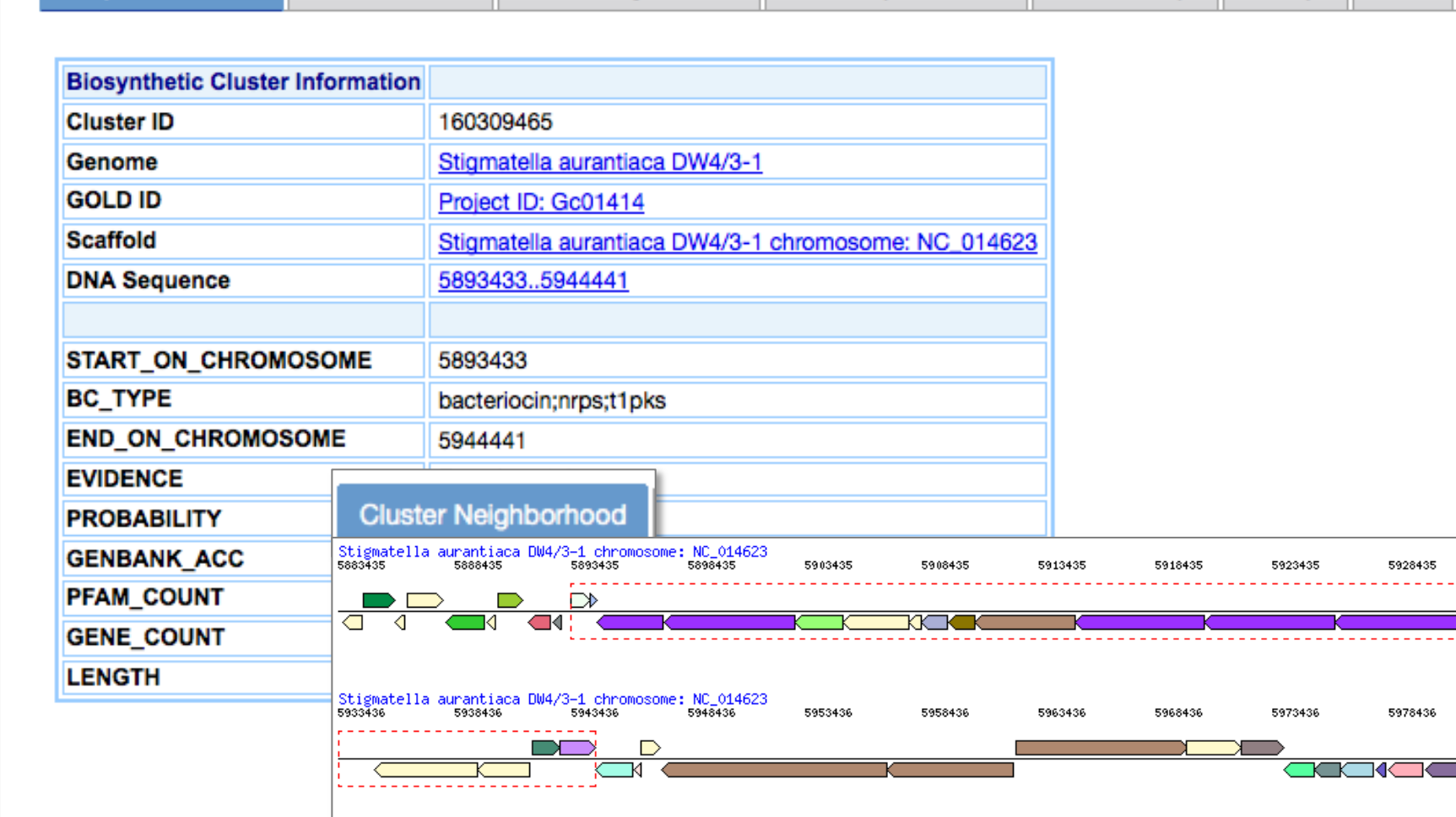


## USER INTERFACE AND ANALYSIS TOOLS

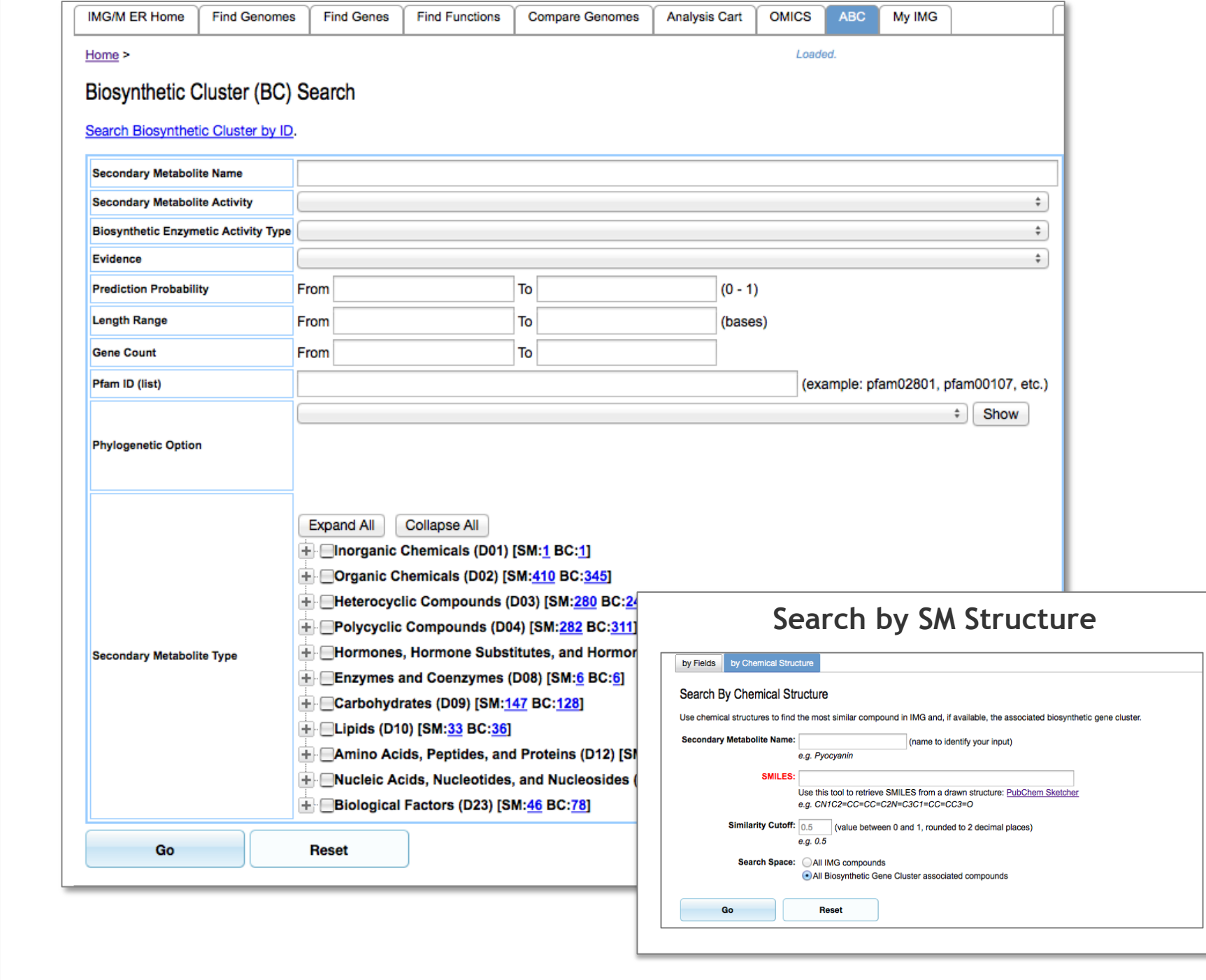
IMG-ABC can be accessed directly at <https://img.jgi.doe.gov/abc/> or through the ABC tab in the IMG system.



## BC Detail Page

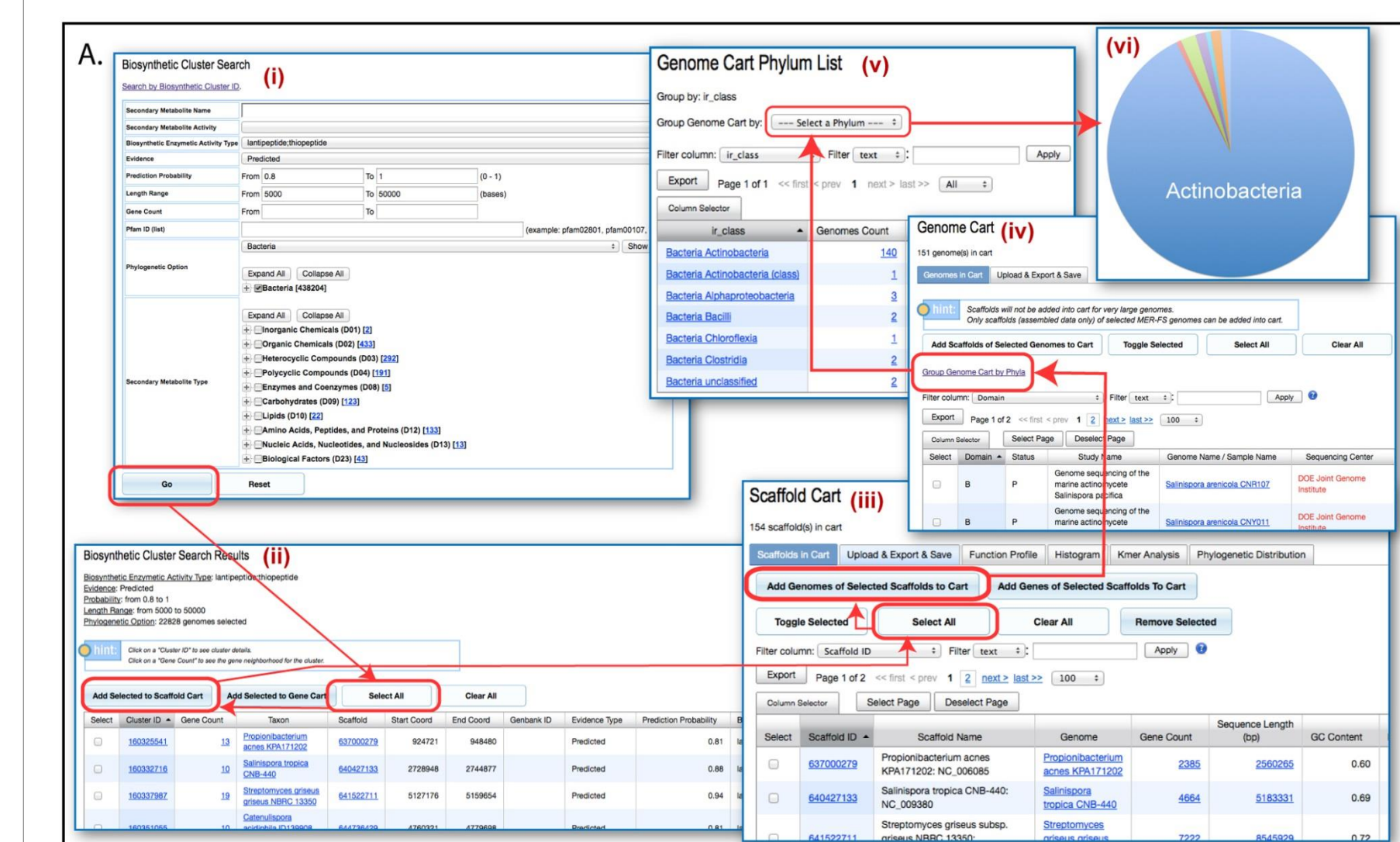


## Search by BC and SM Attributes

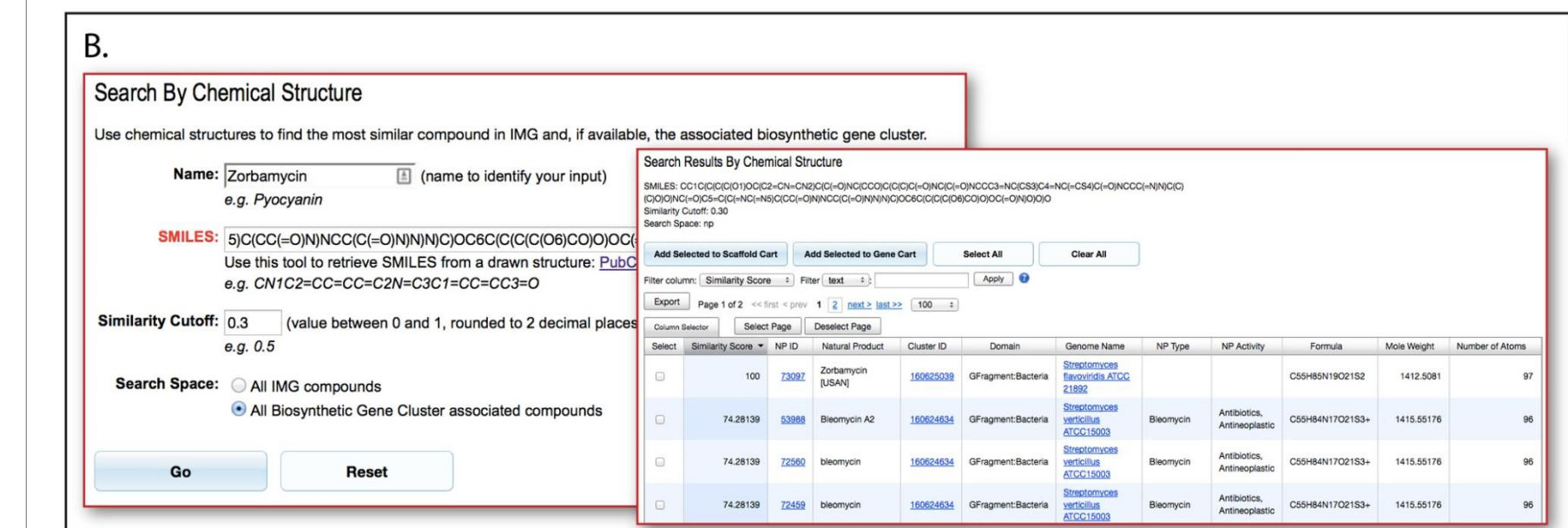


## ANALYSIS WORKFLOW EXAMPLE

(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.



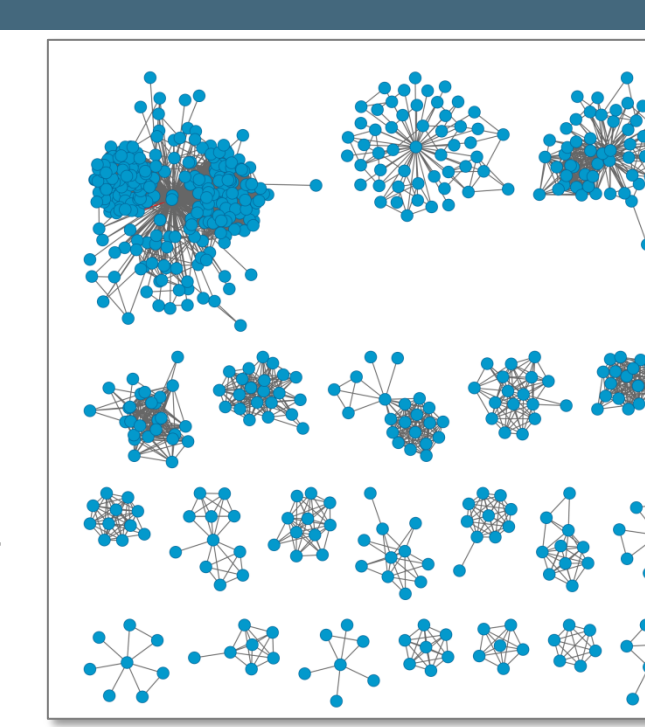
(B) An example of a search using a chemical structure as a query. The chemical structure descriptor string for Zorbamycin is used to search IMG-ABC for secondary metabolites (SM) that match the query with a similarity score greater than 0.3. This yields a list of 118 compounds sorted by descending similarity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

- Markowitz VM, Chen I-M, Chu K, Szeto E, Palaniappan K, Pillay M, Ratner A, Huang J, Pagani I, Tringe S, et al. 2014. IMG/M 4.0 version of the integrated metagenome comparative analysis system. Nucleic Acids Res. 42:D568-D573.
- Cimermančić 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.
- 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.

## ACKNOWLEDGMENTS

This work was supported by the Office of Biological and Environmental Research, Life Sciences Division, U.S. Department of Energy (Contract No. [DE-AC02-05CH1231]). 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-AC02-05CH1231]). 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](mailto:apati@lbl.gov)  
Michalis Hadjithomas, Ph.D. [michalis@lbl.gov](mailto:michalis@lbl.gov)