

# antiSMASH workshop

**Kristin Labby and Marc Chevrette**

Tiny Earth Symposium  
June 12, 2020

# Workshop Goals, Outline, and Resources

## Goals:

1. to be able to use antiSMASH to identify biosynthetic gene clusters
2. interpret these results
3. leave with ideas to use antiSMASH in your TE course or research

## Outline:

Introductory presentation (20 minutes)

Workshop -use antiSMASH in small groups in breakout rooms (20 mins)

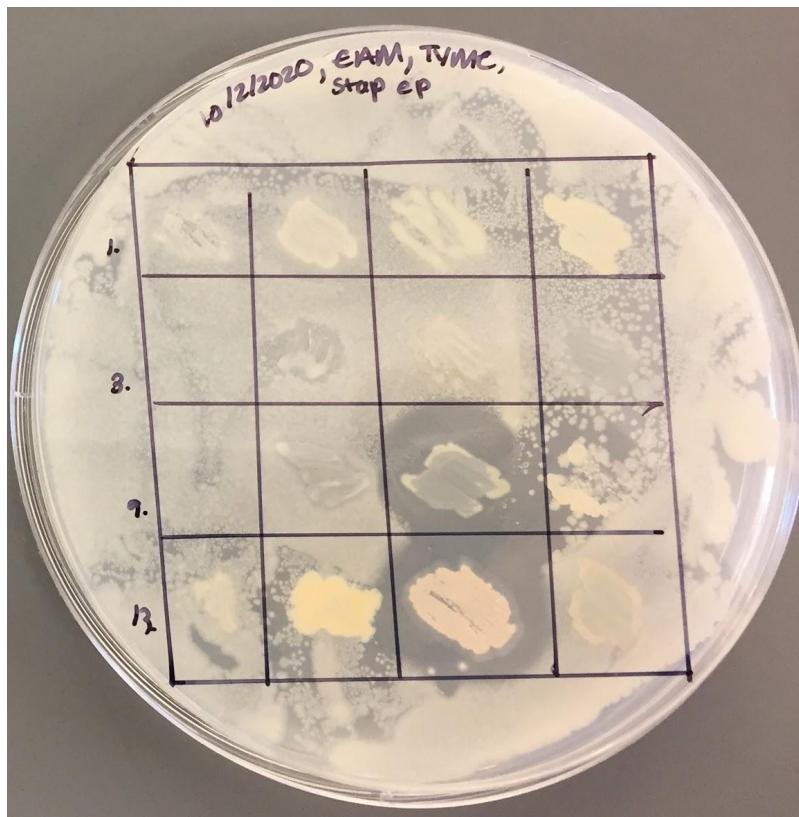
Discuss results (40 minutes)

Wrap up (5 mins)

## Resources:

eHandout, [https://github.com/chevrm/antiSMASH\\_tutorial](https://github.com/chevrm/antiSMASH_tutorial), includes instructions, genome files, and these slides.

# Connecting Tiny Earth to antiSMASH



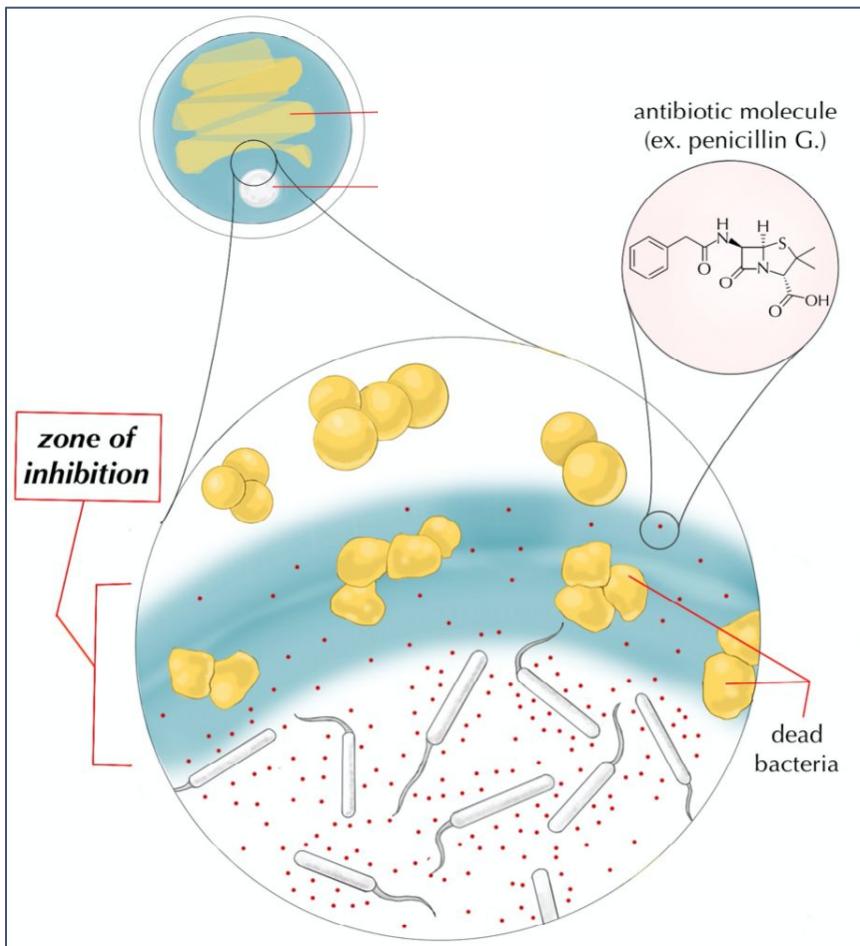
**TE lab goal:** to discover novel, antibiotic-producing bacteria

See Research Guide Section 9, “It all comes down to chemistry”

**zone of inhibition** indicates that a bacterial isolate is producing a *molecule* that has antibiotic properties.

That molecule is a **secondary metabolite**, synthesized from a series of **enzymes** encoded in the isolate's **genome**.

# Connecting Tiny Earth to antiSMASH



**TE lab goal:** to discover novel, antibiotic-producing bacteria

See Research Guide Section 9, “It all comes down to chemistry”

The genes for these biosynthesis enzymes are often spatially clustered together within the genome, referred to as a “**secondary metabolite biosynthesis gene cluster**”.

Due to the similar nature of many of these biosynthesis genes, there is conservation and therefore possibility for **recognition** of these BGCs within a genome.

# Introductory Presentation

## Biosynthesis

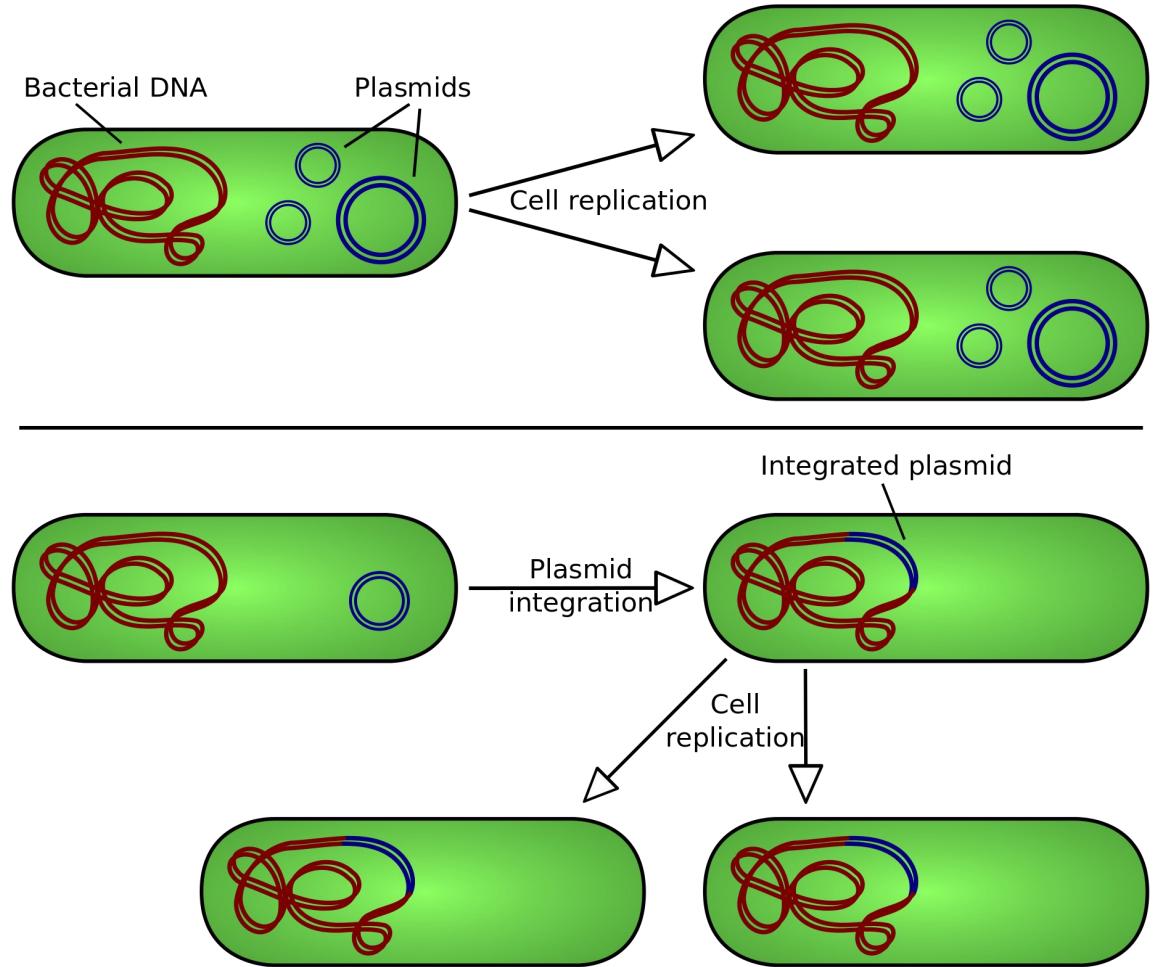
*synthesis of natural products by biological organisms*

## and Biosynthetic Gene Clusters

*genes that encode for enzymes that synthesize natural products*

# the bacterial genome

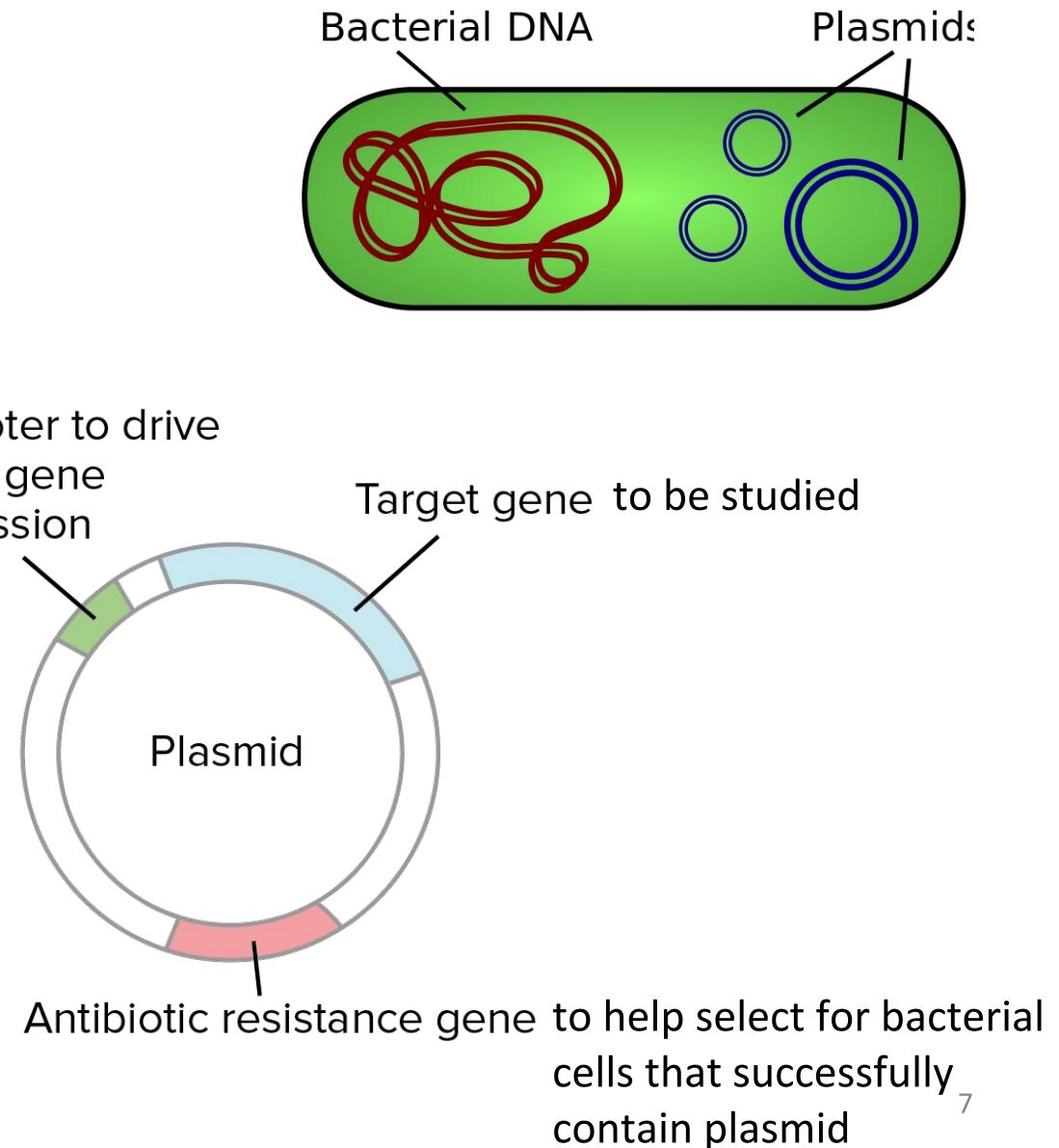
- genome = an organism's entire genetic material
- (Human genome has 23 chromosomes; 3,200 Mbp)
- bacterial genomes are approximately 1-10 million base pairs (1-10 Mbp)
- bacteria also contain genetic info on plasmids, which are "mobile genetic elements"



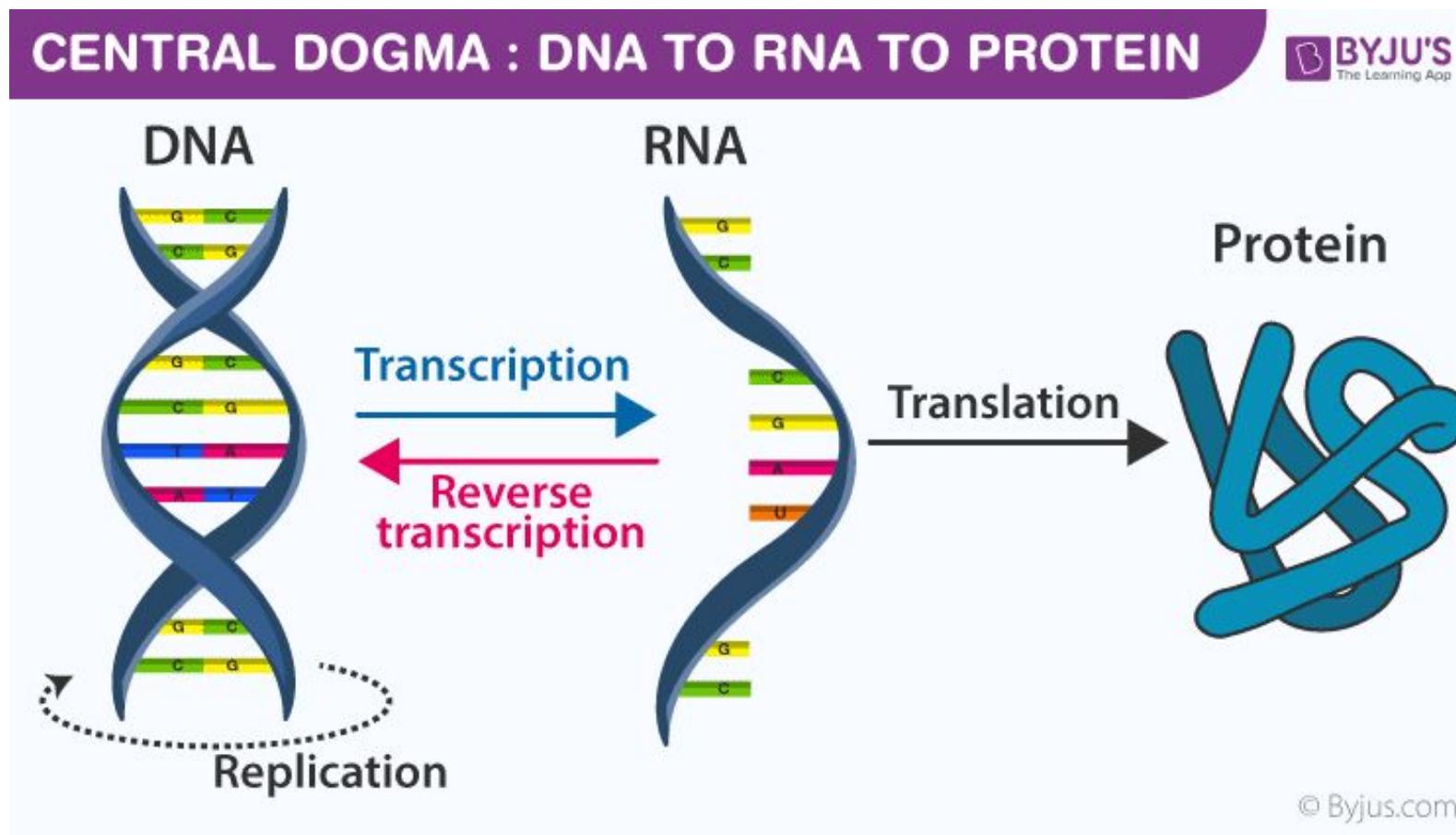
[www.Wikipedia.org/wiki/plasmid](http://www.Wikipedia.org/wiki/plasmid)

# the bacterial genome

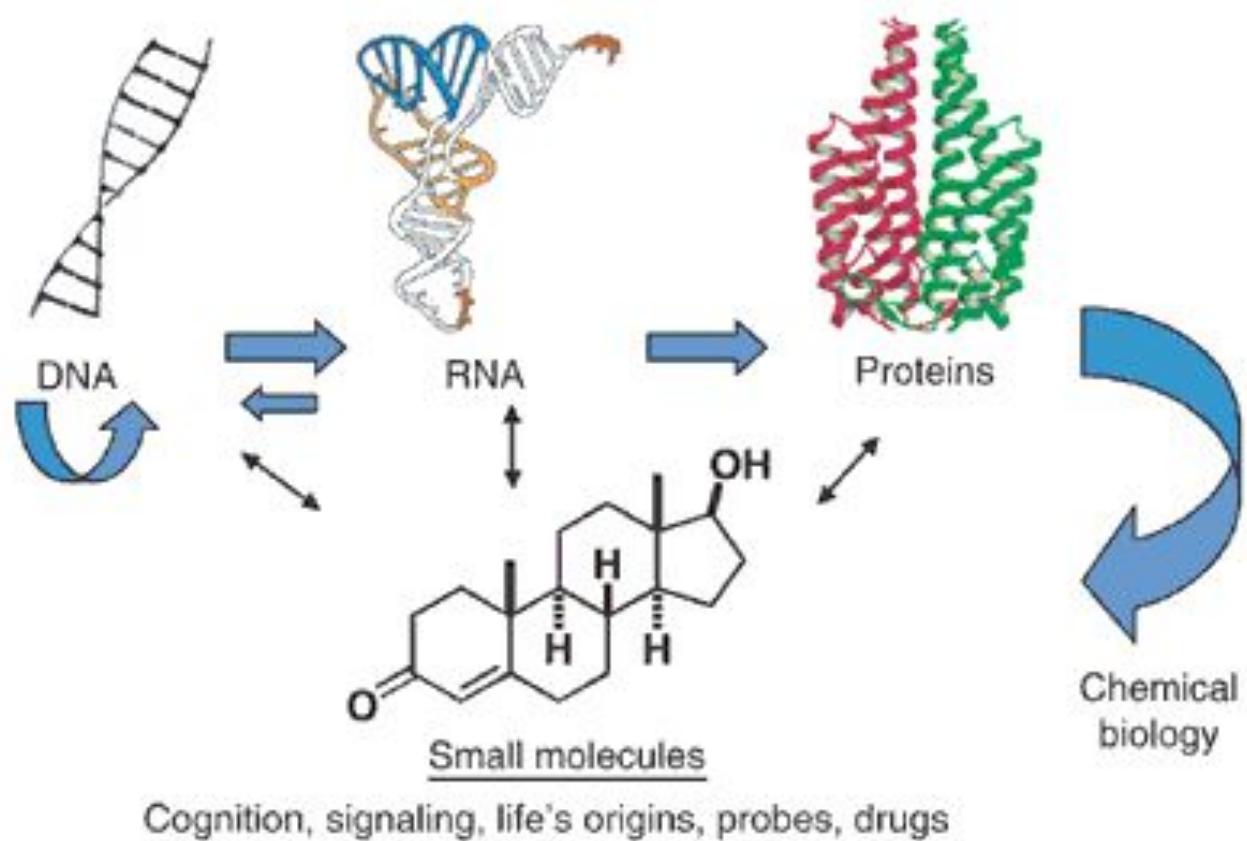
- **plasmids** are easily transferred to other bacteria
- plasmids can contain resistance genes => spread of antibiotic resistance
- bacteria become great organisms to work with in the lab because scientists can easily transform a gene into “competent” bacterial cells (most often *E.coli*)
  - competent means capable of taking up DNA



the central dogma: proteins are the “active agents” for cellular function

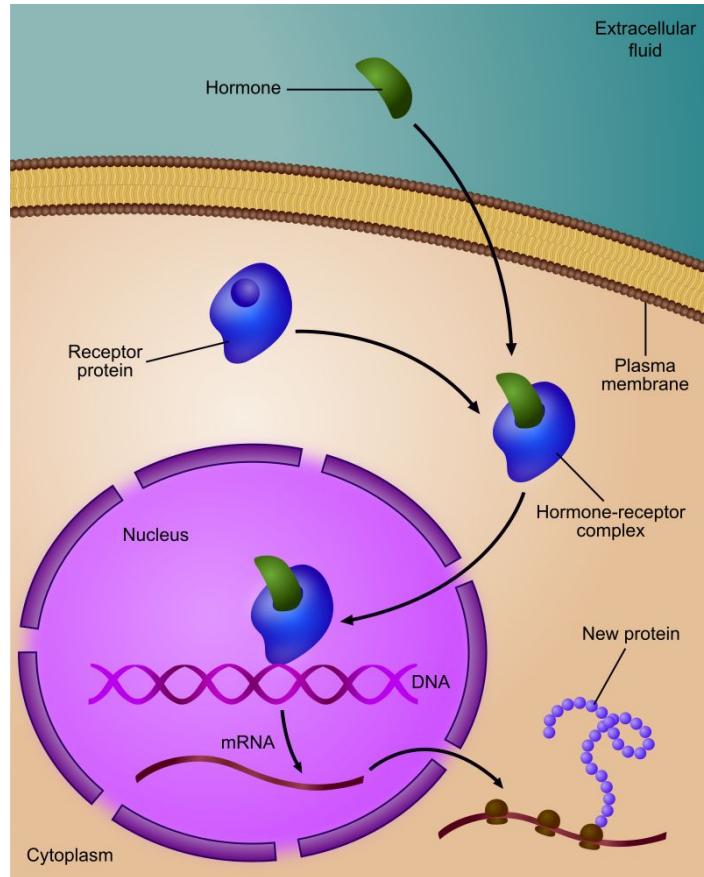


# the central dogma: small molecules play important roles in regulation at each step

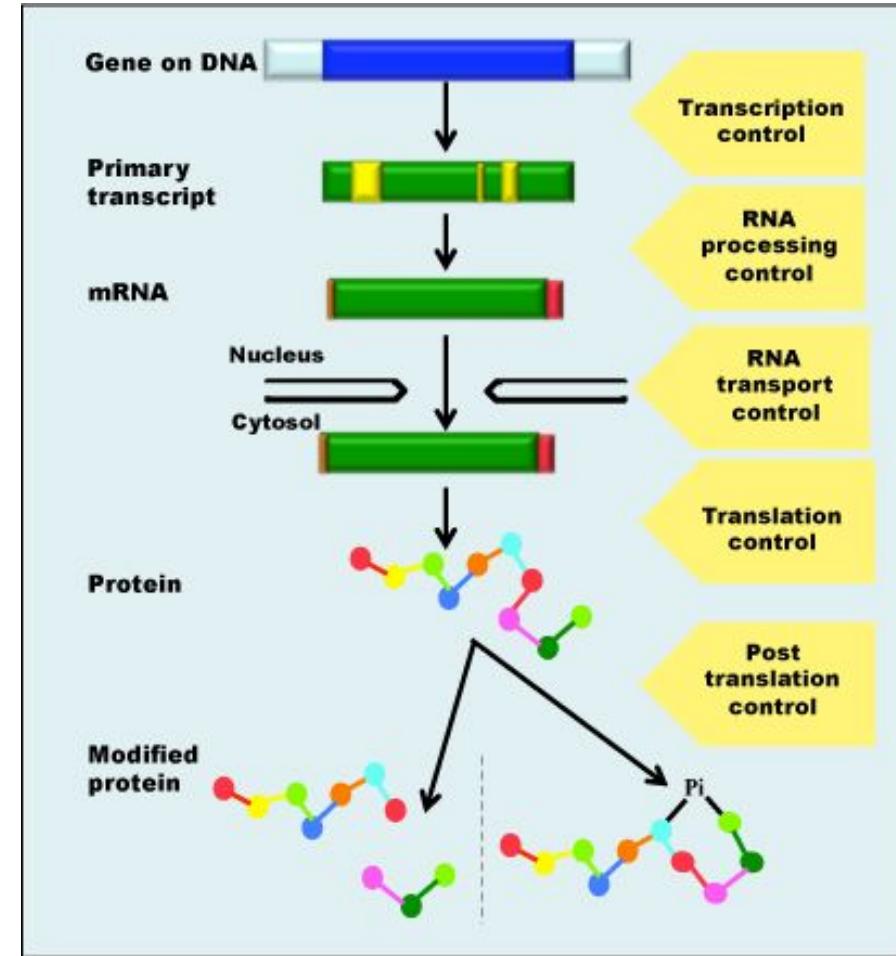


- also, complex small molecules (natural products) are synthesized by proteins (which are encoded by genes)
- about 5-25% of bacterial genome is dedicated to production of secondary metabolites

# gene regulation: not all genes are “turned on” and actively expressing proteins all the time



By Ali Zifan 03:07, 10 July 2016 (UTC) - Own work; Used information from: Campbell Biology (10th Edition) by: Jane B. Reece & Steven A. Wasserman., CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=50049531>

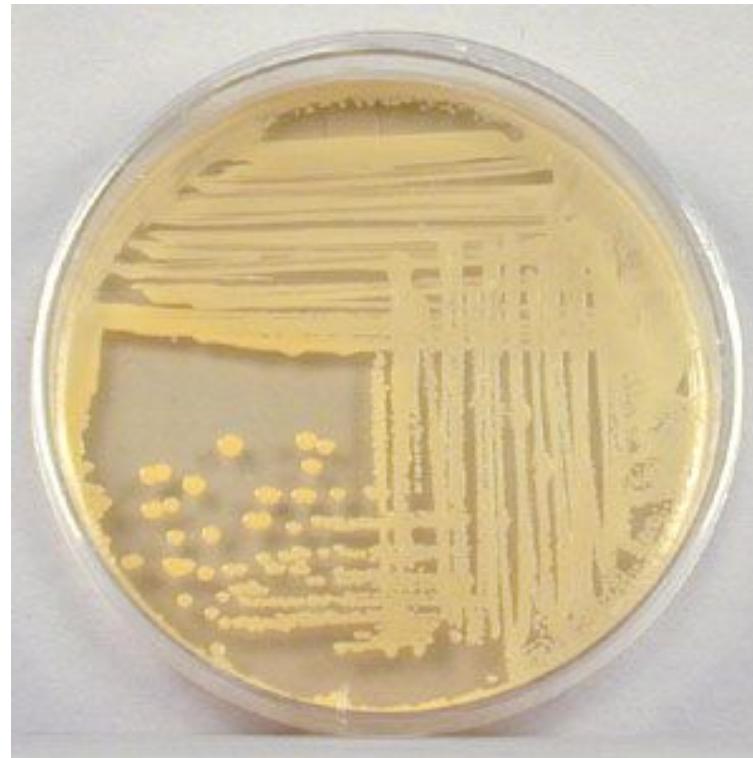


OpenStax Human Physiology Modified from: ArneLH, Wikimedia commons, Gene\_expression\_control.png

Both photos are *Janthinobacterium lividum*...  
What's going on? Why are they different?



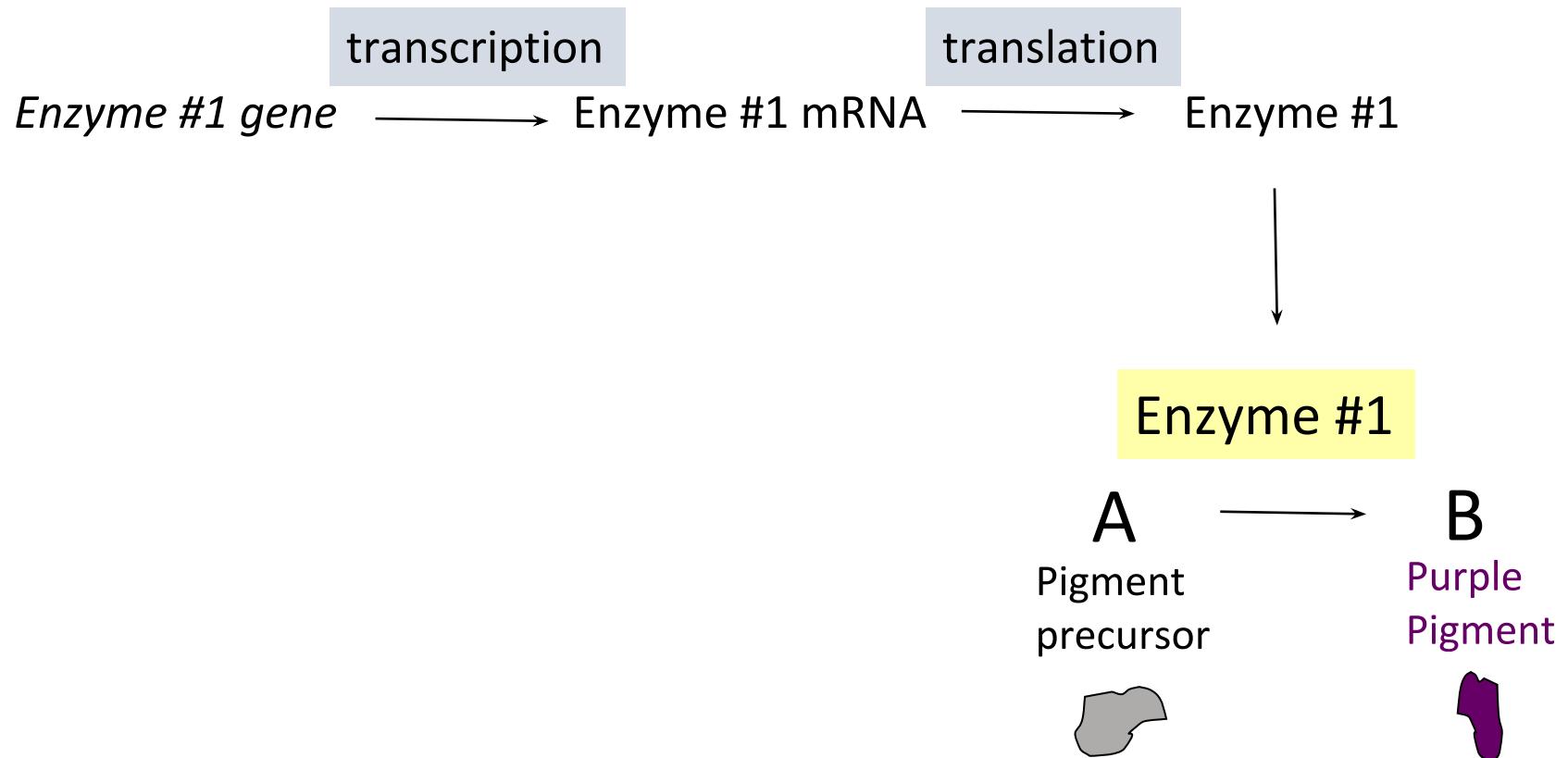
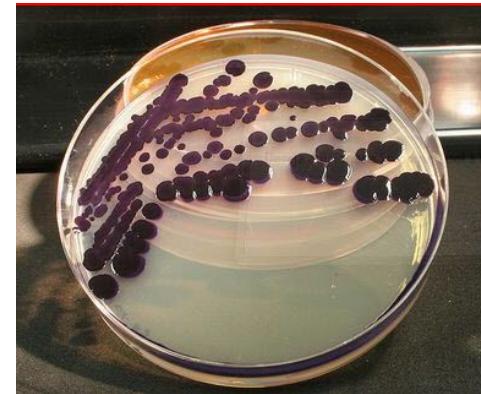
<http://microbiologiabrasil.blogspot.com/2009/01/janthinobacteriumsp.html>



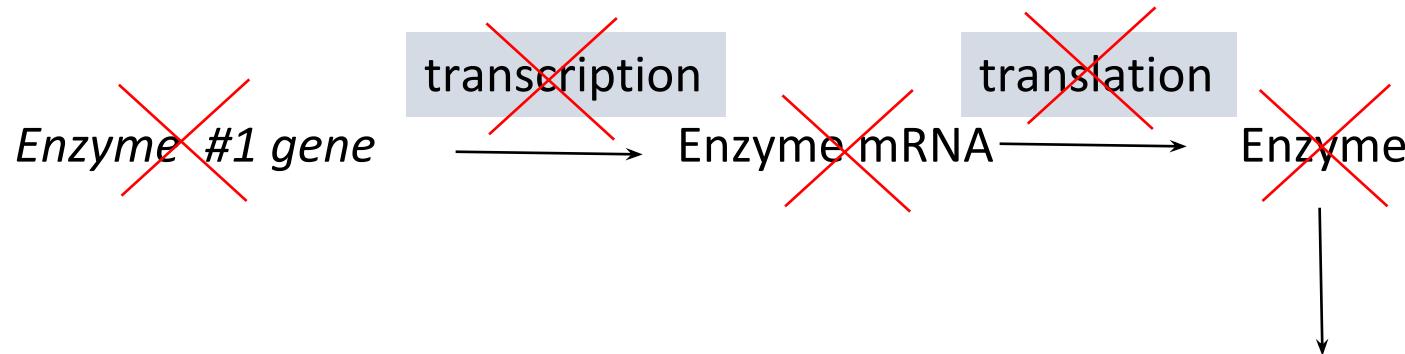
Copyright © Gary E. Kaiser

<http://faculty.ccfcmd.edu/courses/bio141/labmanual/lab2/sainsol.html>

Here's a clue... there is a specific enzyme; lets call it "Enzyme #1" that does the final step in synthesis of the purple pigment

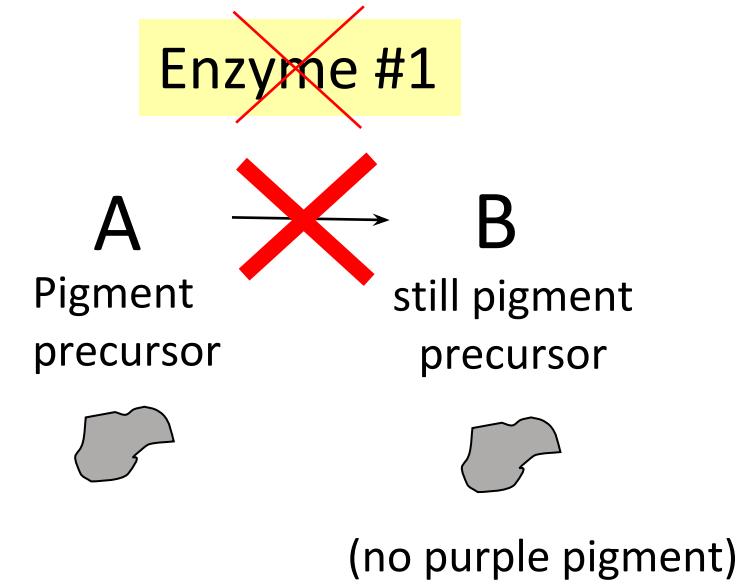


One possibility: the gene for Enzyme #1 is missing from this organism.

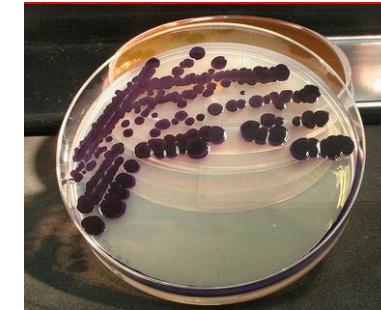
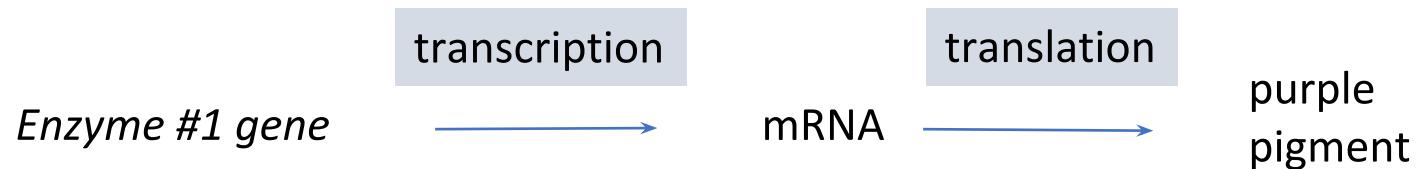


Gene for enzyme missing:

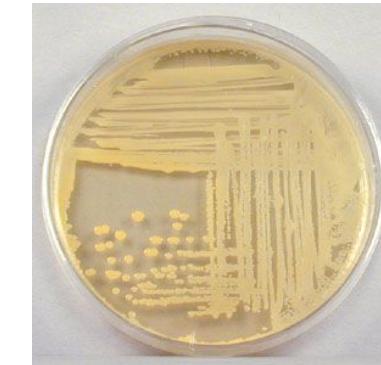
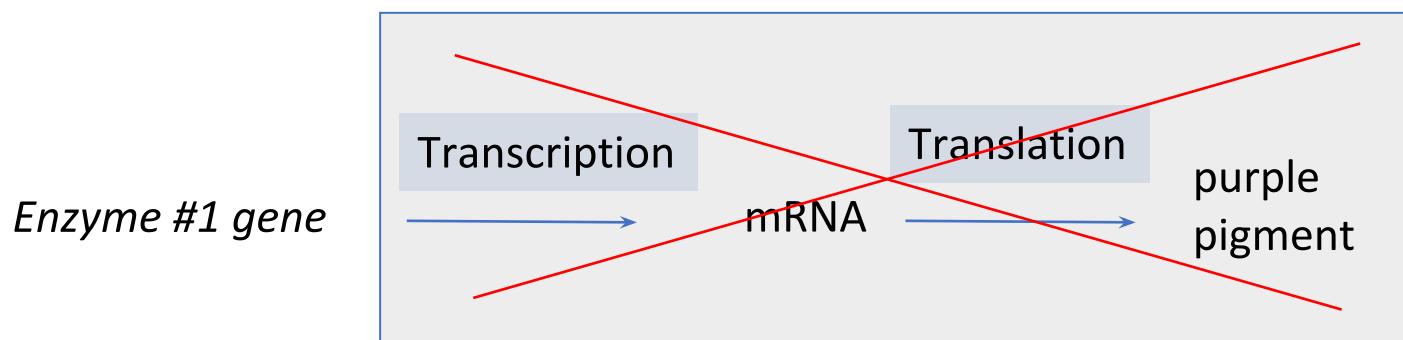
- No transcription of enzyme
- No translation of enzyme
- No enzyme protein
- No purple pigment accumulates



# Another possibility: the gene for Enzyme #1 is turned off.



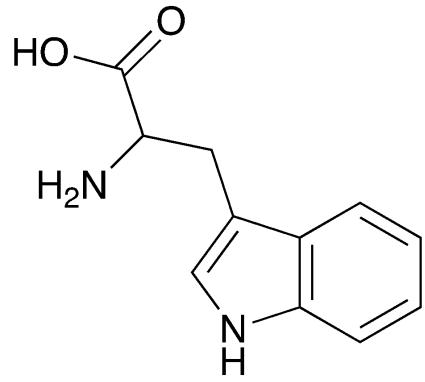
Purple colony



White colony

The purple colony is genetically identical to the white colony; just not expressing some or all of the genes necessary for purple pigment biosynthesis.

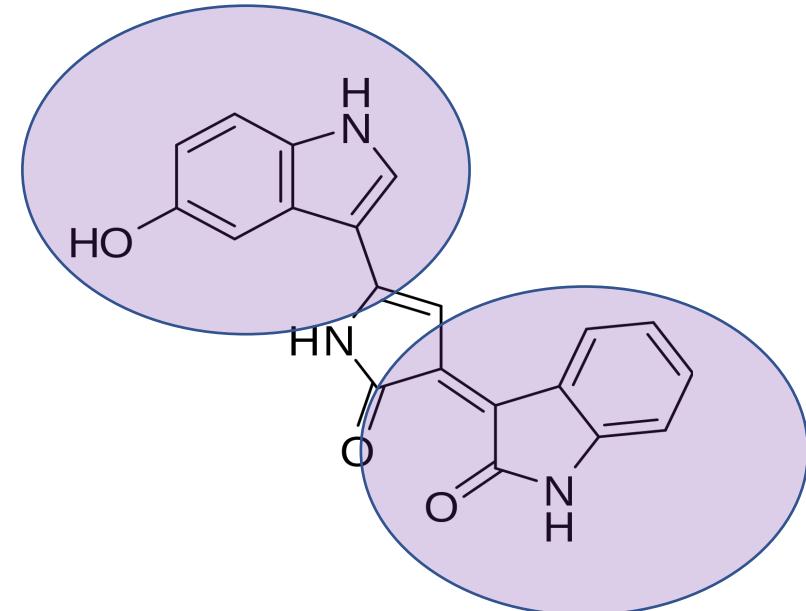
**purple pigment** is synthesized by many enzymes from many genes: a **biosynthetic gene cluster**.



tryptophan



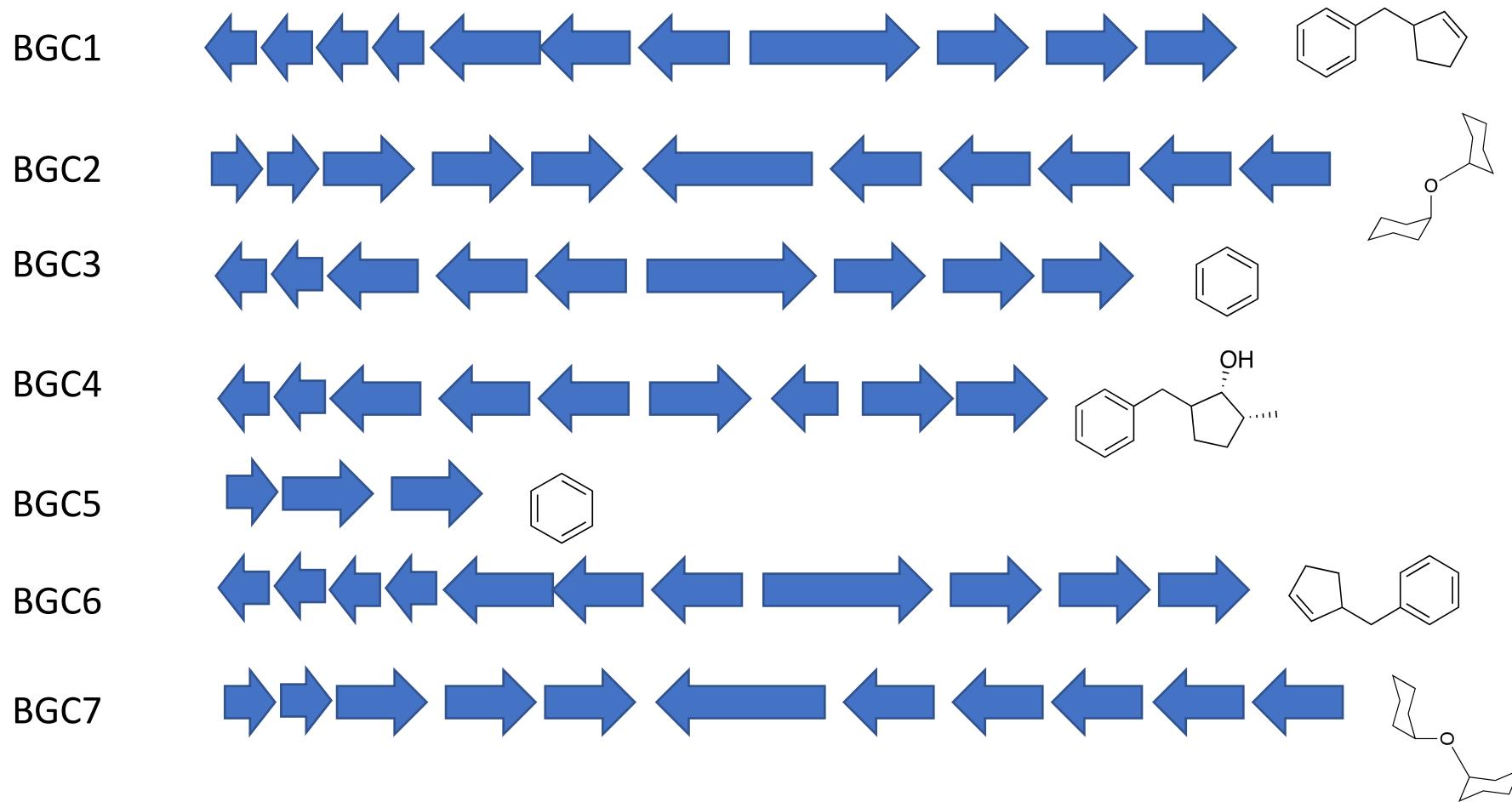
in our scenario; “Enzyme 1” could be any of these enzymes in the pathway for violacein synthesis



violacein  
(purple pigment)

two molecules of tryptophan serve as precursors for violacein

bacterial genome typically contains 10 – 40 BGCs

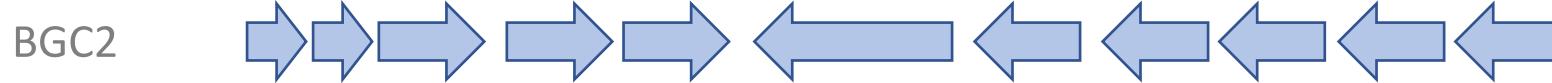


**cryptic or silent BGCs:**  
not all BGCs are expressed at the same time; in fact expression can change under different conditions.

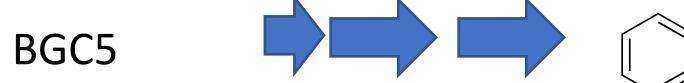
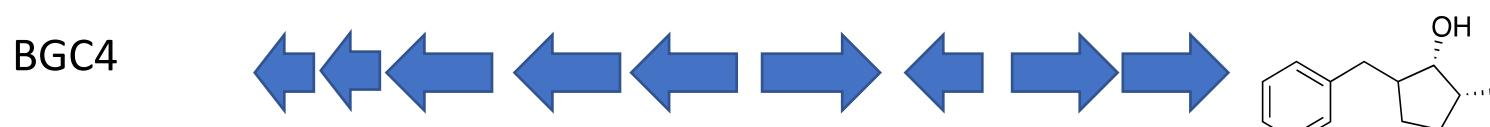
**cryptic or silent BGCs:** not all BGCs are expressed at the same time; in fact expression can change under different conditions.



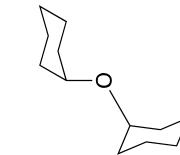
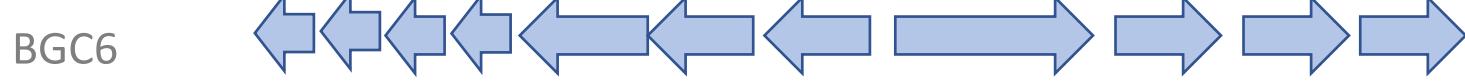
example:  
“condition 1”



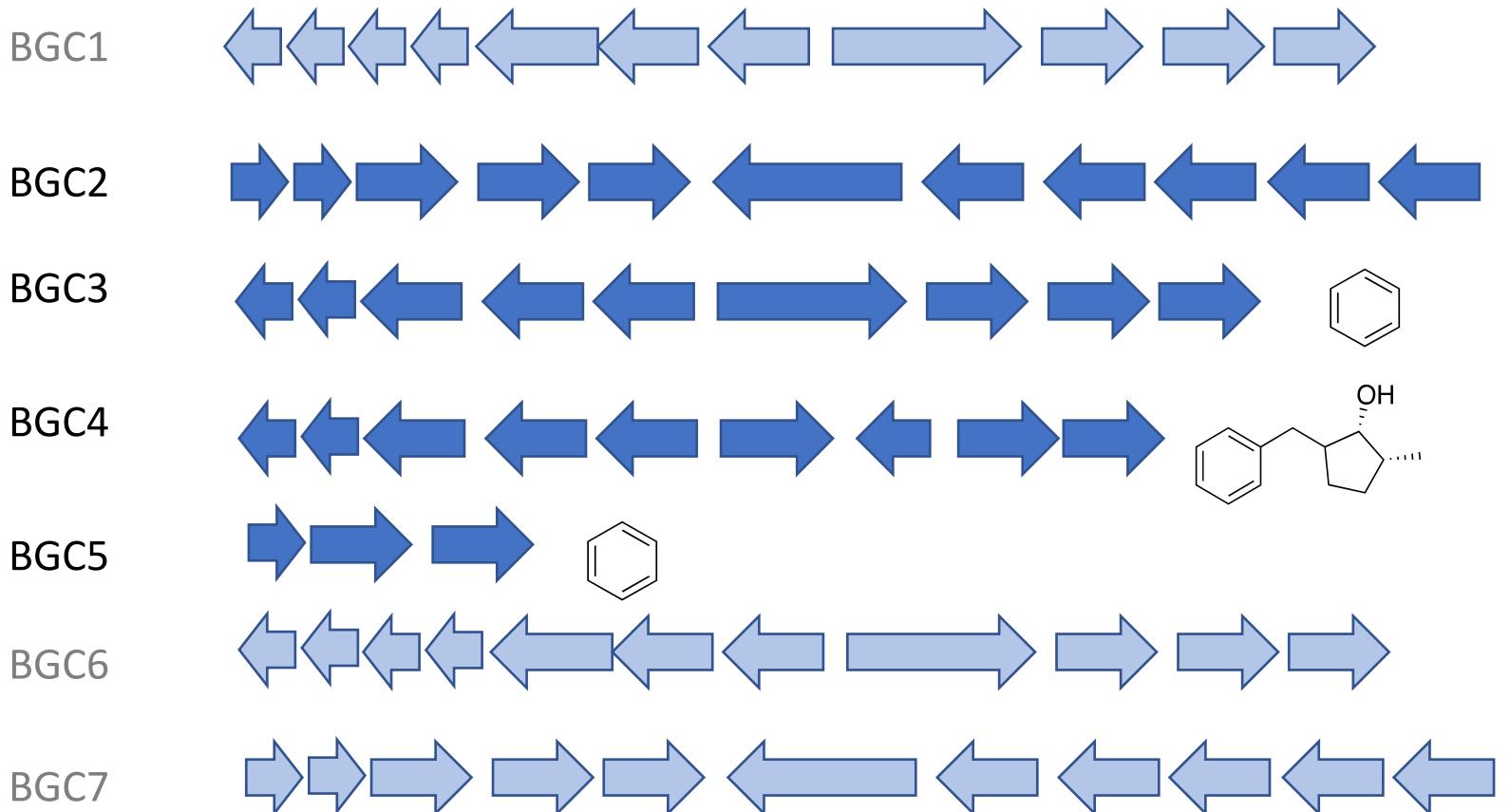
BGC1, 4, 5 and 7  
are expressed



BGC2, 3 and 6  
are repressed



**cryptic or silent BGCs:** not all BGCs are expressed at the same time; in fact expression can change under different conditions.



example:  
“condition 2”

BGC2, 3, 4, and 5  
are expressed

BGC1, 6 and 7  
are repressed

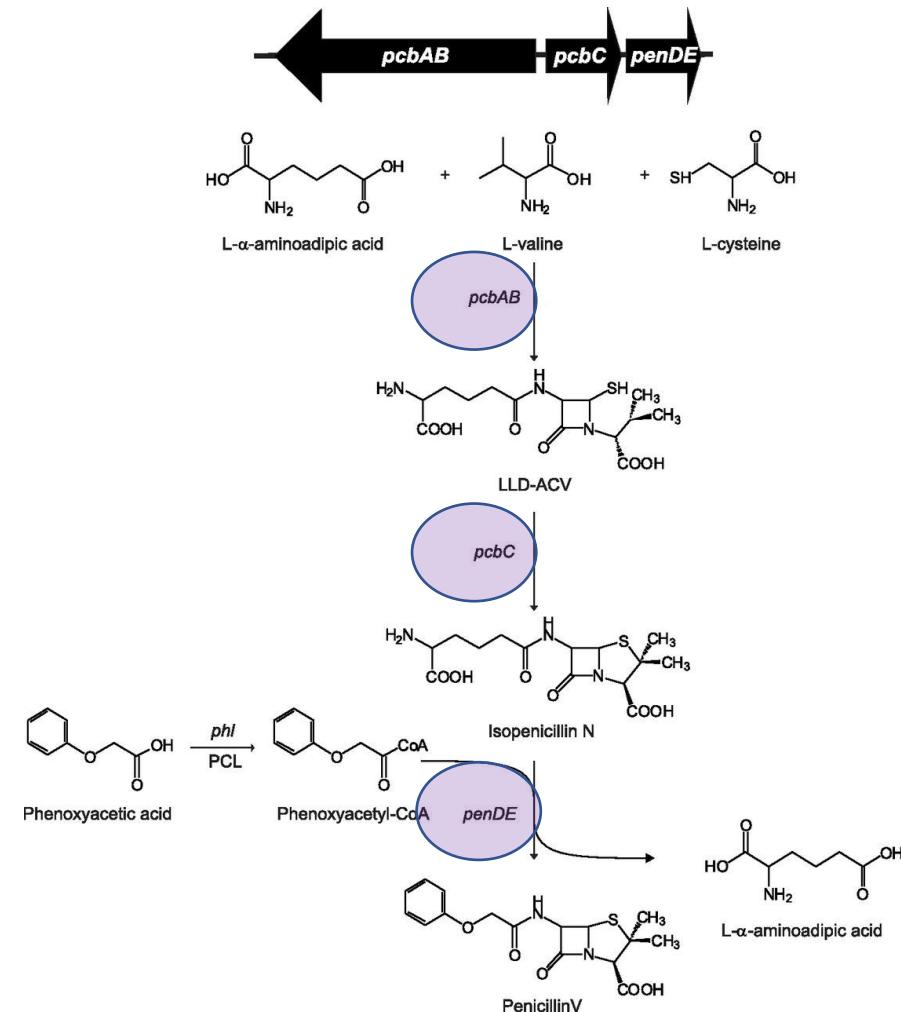
# Sometimes biosynthetic gene clusters are simple....

penicillin V

synthesized by

*Penicillium chrysogenum*  
(fungus)

(many other organisms also  
synthesize penicillins!)



Increased Penicillin Production in *Penicillium chrysogenum* Production Strains via Balanced Overexpression of Isopenicillin N Acyltransferase

Stefan S. Weber, Fabiola Polli, Rémon Boer, Roel A. L. Bovenberg, Arnold J. M. Driessens

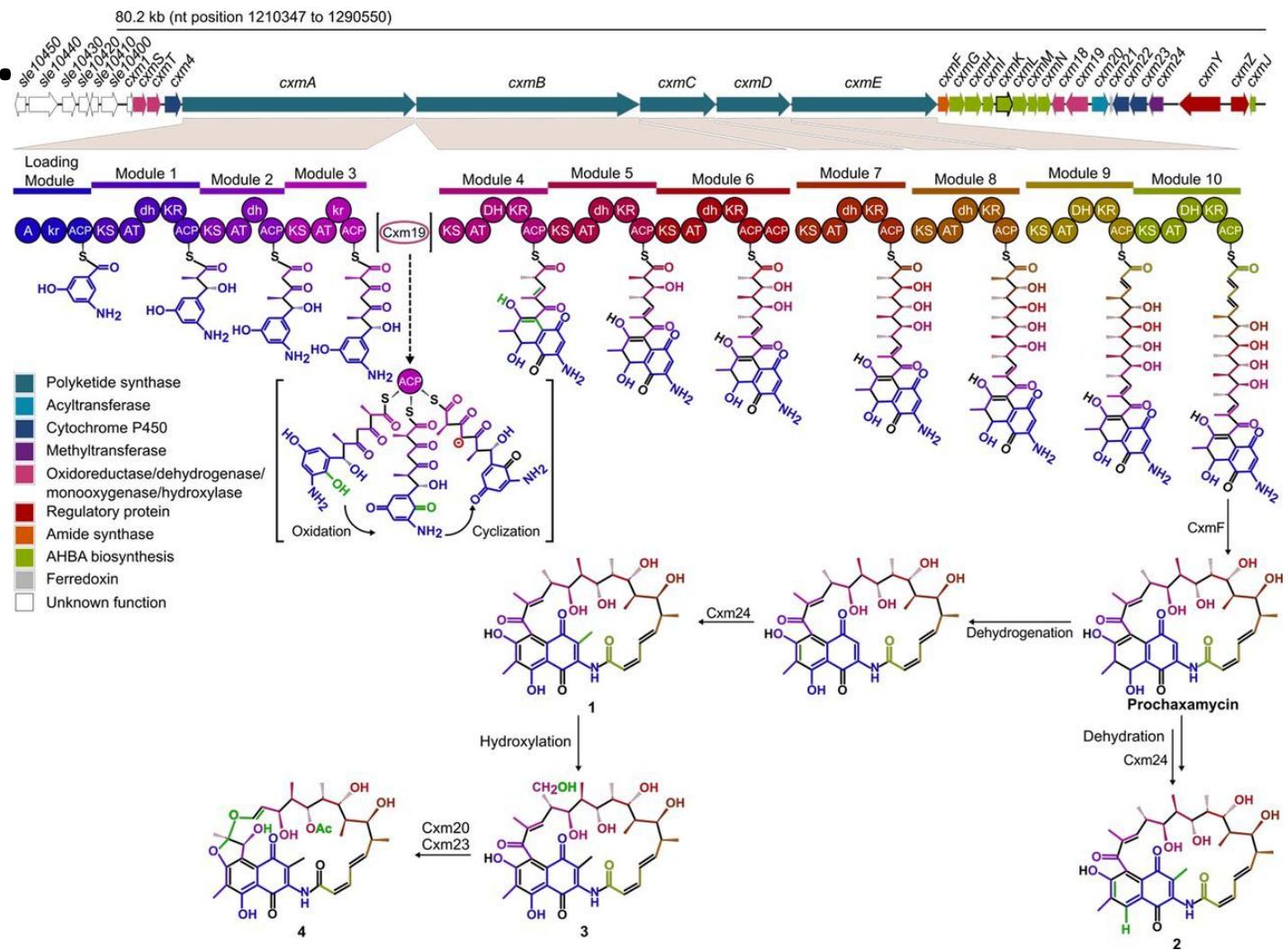
Applied and Environmental Microbiology Sep 2012, 78 (19) 7107-7113; DOI: 10.1128/AEM.01529-12

# Sometimes biosynthetic gene clusters are really complex....

chaxamycins A to D

synthesized by

*Streptomyces leeuwenhoekii*



Identification and Heterologous Expression of the Chaxamycin Biosynthesis Gene Cluster from *Streptomyces leeuwenhoekii*

Jean Franco Castro, Valeria Razmilic, Juan Pablo Gomez-Escribano, Barbara Andrews, Juan A. Asenjo, Mervyn J. Bibb

Applied and Environmental Microbiology Aug 2015, 81 (17) 5820-5831; DOI: 10.1128/AEM.01039-15

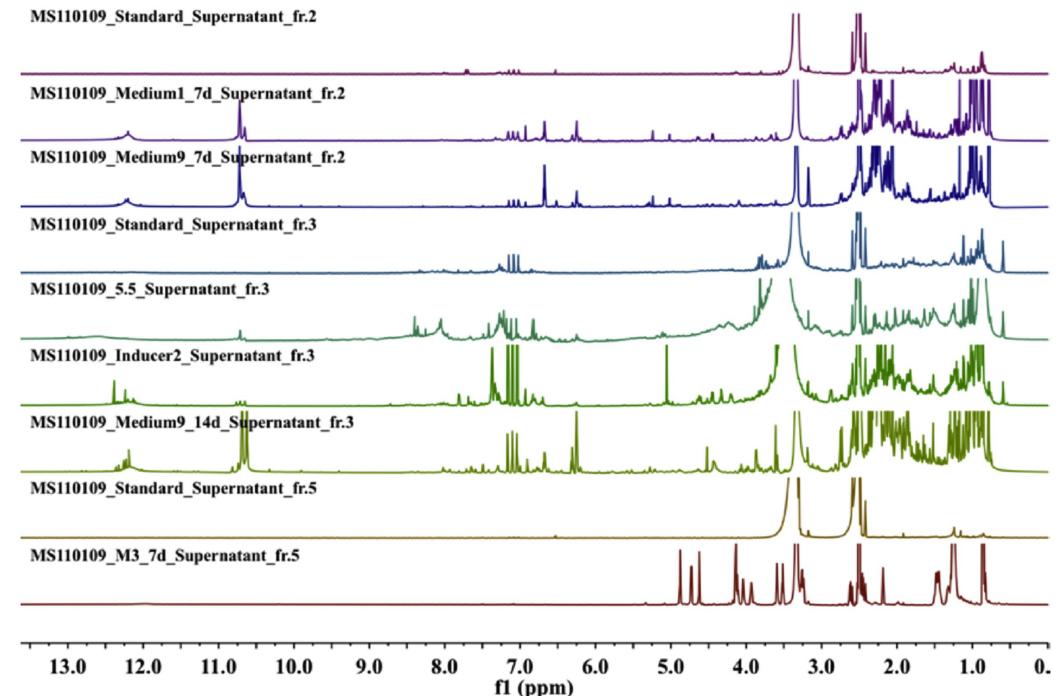
# How do scientists learn about biosynthetic gene clusters?

- which organisms contain which BGCs?
- when are they turned on in that organism?
- what compound(s) do(es) the BGC make?

# Traditional, laboratory approach to learn about BGCs: expression can change under different conditions: OSMAC “One Strain Many Compounds”



Figure 2 Photos of G2 in liquid media (top): from left to right: Czapek dox. 2% malt extract. potato dextrose. YPSS. YESD. and PYG



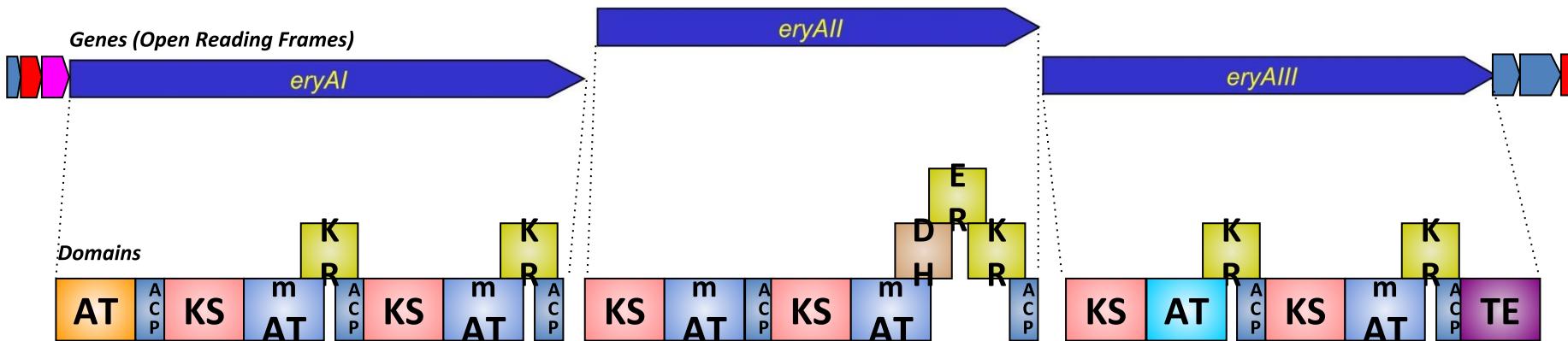
Evaluation of culture media for the production of secondary metabolites in a natural products screening program

Karen M Vandermolen, Huzefa A. Raja, Tamam El-Elimat, Nicholas H Oberlies. Published in AMB Express 2013

A systems approach using OSMAC, Log P and NMR fingerprinting: An approach to novelty. Synthetic and Systems Biotechnology  
Volume 2, Issue 4, December 2017, Pages 276-286

# Modern, genomic method to learn about BGCs: genome-mining and programs like antiSMASH

- within the past 5 years the ability to sequence entire bacterial genome has become routine
- >300,000 bacterial genomes have been deposited into NCBI database
- many BGCs have been characterized and published
  - BGCs are conserved across types of natural products



*Colinearity:* Order of modules and genes often corresponds with order of synthesis

Erythromycin is awesome!

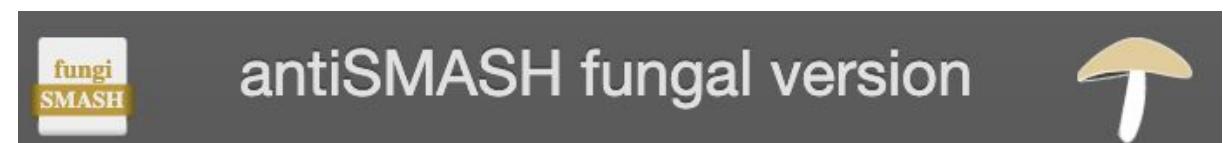
*"Well then it should be really easy to predict - us."*

*"Not so fast, everyone..."*  
-Nature

# Modern, genomic method to learn about BGCs: genome-mining and programs like **antiSMASH**

**antiSMASH** = a program that enables genomic analysis for enzyme signatures to predict biosynthetic potential of an organism

## antibiotics & Secondary Metabolite Analysis Shell



# Modern, genomic method to learn about BGCs: genome-mining and programs like antiSMASH

Relies on submitted annotated BGCs from MIBiG:

The screenshot shows the MIBiG website interface. At the top, there's a navigation bar with links for Home, Statistics, Search, Repository, Submit, and Download. Below the header, the MIBiG logo is displayed with the text "Minimum Information about a Biosynthetic Gene cluster". A central diagram illustrates the workflow: "Submitting scientists" provide data to the "MIBiG Repository", which then feeds into "Multiple online databases / services" like "antiSMASH" and "JGI". These databases are shown with icons of servers and databases. The bottom section of the diagram is divided into three numbered steps: 1. Connecting genes to chemistry (showing a chemical structure), 2. Understanding BGC environmental diversity (showing a world map with red dots), and 3. Computer-guided gene cluster engineering (showing a DNA sequence with arrows). A descriptive text at the bottom states: "The Minimum Information about a Biosynthetic Gene cluster (MIBiG) specification provides a robust community standard for annotations and metadata on biosynthetic gene clusters and their molecular products. MIBiG is a Genomic Standards Consortium project that builds on the Minimum Information about any Sequence (MiSx) framework."

Connects with NCBI databases:

The screenshot shows the NCBI Genome page. At the top, there's a banner with a COVID-19 warning: "COVID-19 is an emerging, rapidly evolving situation. Get the latest public health information from CDC: <https://www.coronavirus.gov>. Get the latest research from NIH: <https://www.nih.gov/coronavirus>". Below the banner, the page title is "Genome". A large image of chromosomes is on the left. The main content area is titled "Using Genome" and includes links for Help, Browse by Organism (UPDATED), Download / FTP, Download FAQ, and Submit a genome. To the right, there are sections for "Custom resources" (Human Genome, Microbes, Organelles, Viruses, Prokaryotic reference genomes), "Other Resources" (Assembly, BioProject, BioSample, Genome Data Viewer NEW), "Genome Tools" (BLAST the Human Genome, Microbial Nucleotide BLAST), "Genome Annotation and Analysis" (Eukaryotic Genome Annotation, Prokaryotic Genome Annotation, PASC (Pairwise Sequence Comparison)), and "External Resources" (GOLD - Genomes Online Database, Bacteria Genomes at Sanger, Ensembl). At the bottom, there are links for "GETTING STARTED", "RESOURCES", "POPULAR", "FEATURED", and "NCBI INFORMATION", along with a "Support Center" link.

# Statistics

## General statistics

Total Secondary Metabolite Clusters:	1923
Minimal entries:	1434
Complete entries:	465
Non-minimal entries:	489
Incomplete entries:	27

## Secondary Metabolite record counts

Polyketide	799	Nonribosomal peptide	605
Ribosomally synthesized and post-translationally modified peptide	258	Other	253
Saccharide	173	Terpene	130
Alkaloid	49		

## Entries

Streptomyces	636	Aspergillus	88
Pseudomonas	68	Bacillus	56
Unclassified	50	Streptococcus	34
Penicillium	28	Burkholderia	27
Micromonospora	26	Amycolatopsis	25

[Home](#)[Using antiSMASH](#)[Introduction](#)[Submitting jobs on the website](#)[Understanding the output](#)[Module-specific information](#)[BGC detection](#)[NRPS/PKS domains](#)[NRPS/PKS modules](#)[PKS type II](#)[Local antiSMASH installation](#)[Installation](#)[Starting command line runs](#)[antiSMASH5 feature list](#)[Change log](#)[5.1](#)[FAQ](#)[Glossary](#)[Cluster Types](#)[Current Types](#)[Obsolete/Previous Types](#)[About](#)

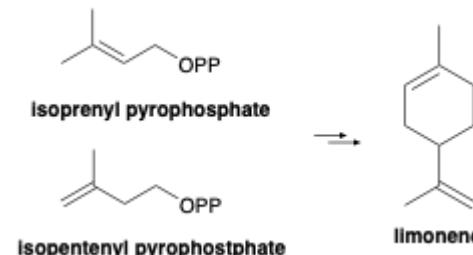
# Cluster Types

antiSMASH uses some abbreviations internally to refer to the different types of secondary metabolite clusters, a short explanation of the different types can be found below:

## Current Types

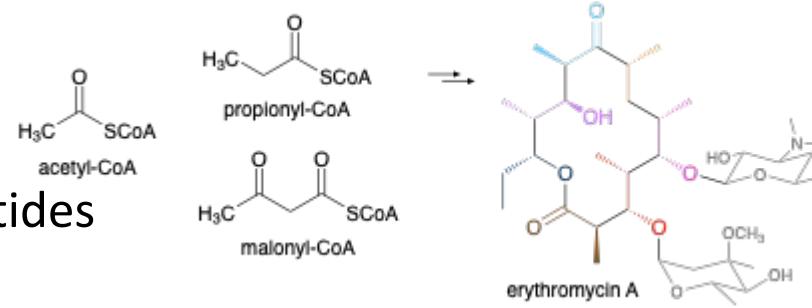
Label	Description	Added	Last updated
T1PKS	Type I PKS (Polyketide synthase)	<= 3.0	<= 3.0
T2PKS	Type II PKS	<= 3.0	5.0
T3PKS	Type III PKS	<= 3.0	<= 3.0
transAT-PKS	Trans-AT PKS	<= 3.0	5.0
transAT-PKS-like	Trans-AT PKS fragment, with trans-AT domain not found	<= 5.0	5.0
PpyS-KS	PPY-like pyrone cluster	4.2	4.2
hglE-KS	heterocyst glycolipid synthase-like PKS	5.0	5.0
CDPS	tRNA-dependent cyclodipeptide synthases	5.0	5.0
PKS-like	Other types of PKS cluster	5.0	5.0
arylpolyene	Aryl polyene cluster	<= 3.0	<= 3.0
resorcinol	Resorcinol cluster	<= 3.0	<= 3.0
ladderane	Ladderane cluster	<= 3.0	<= 3.0
PUFA	Polyunsaturated fatty acid cluster	<= 3.0	<= 3.0
nrps	Non-ribosomal peptide synthetase cluster	<= 3.0	<= 3.0
nrps-like	NRPS-like fragment	5.0	5.0
thioamide-NRP	Thioamide-containing non-ribosomal peptide	5.0	5.0

# there are 7 major classes of natural products



## 1. polyketides

- a. aromatic polyketides
- b. macrolactones
- c. decalin-containing polyketides
- d. polyenes
- e. polyethers



## 2. non-ribosomal polypeptides

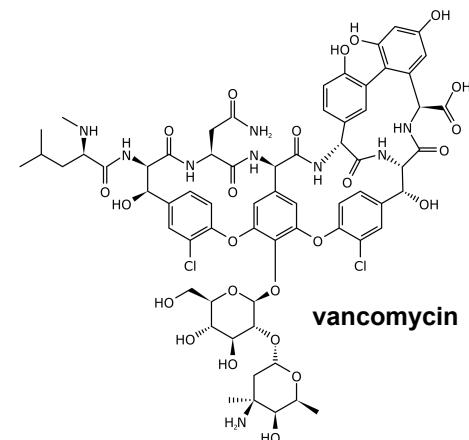
## 3. isoprenoids/terpenes

## 4. alkaloids

## 5. glycosides (saccharides)

## 6. nucleosides

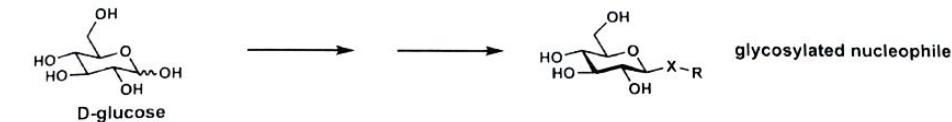
## 7. phenylpropanoids



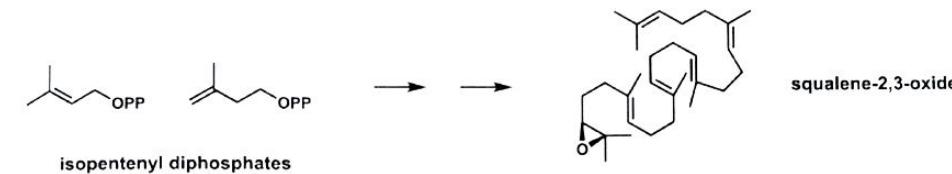
- polyketides and non-ribosomal polypeptides are built on enzymatic assembly lines (called polyketide synthetases (PKSs) and non-ribosomal peptide synthetases (NRPSs))
- most *therapeutically interesting* molecules are built from hybrid NRP-PK synthetases (ex. rapamycin, FK506, bleomycin, epothilones)
- many natural products (of all classes) are glycosylated during final synthetic steps

primary metabolites serve as building blocks for specific classes of natural products

glycosides



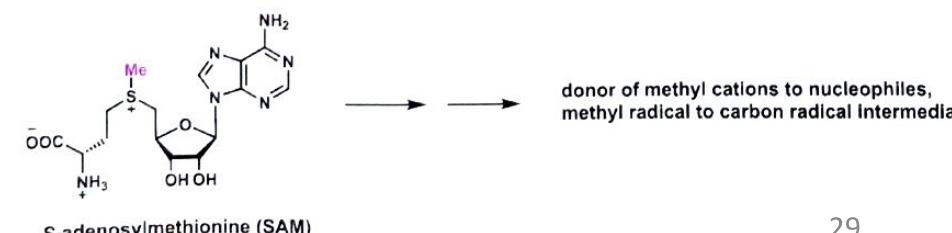
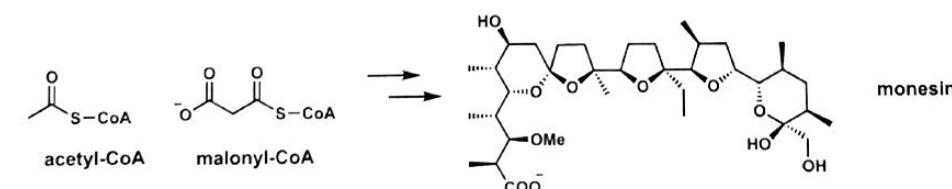
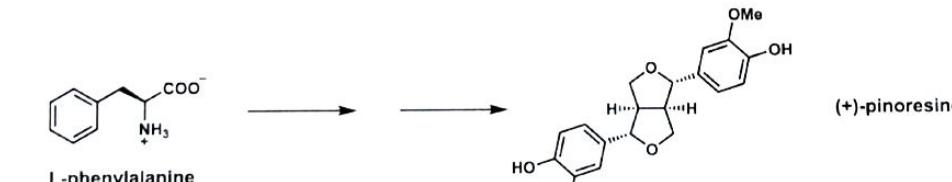
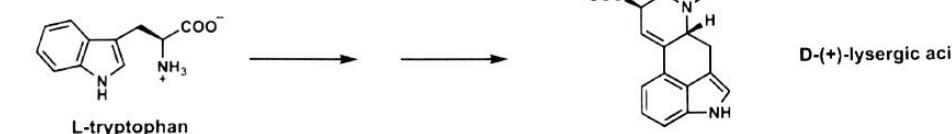
isoprenoids/  
terpenes



alkaloids  
(N-containing)

non-ribosomal  
polypeptides

polyketides



# Using antiSMASH in the context of Tiny Earth

## 1. One of our TE isolates.

- a. If available, use the complete genomic nucleotide sequence of one of our isolates, or use a genomic TE sequence file shared from the ChemHub.
- b. If you had successful PCR and sequencing of the 16S rRNA gene of one of your isolates, you can use the genomic, nucleotide sequence of the top BLAST match.

## 2. An interesting bacterial secondary metabolite.

Use the list of natural products below to use antiSMASH to identify the biosynthetic gene cluster responsible for making that bacterial secondary metabolite.

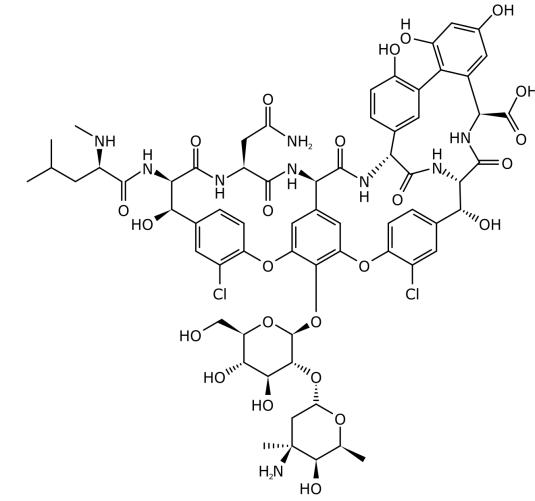
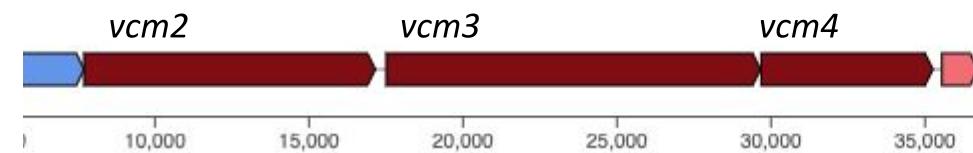
- a. Table 1, below, is a list of interesting natural products. Your job will be to do some research to find what bacterial organism(s) make that natural product. Then search NCBI [genome database](#) to find the genomic, nucleotide accession number for that species.  
(You may also add your own bacterial natural product of interest to Table 1!)
- b. Use this genome to run antiSMASH. Please add the species name and accession number to Table 1.

Bacterial species → genome contains biosynthetic gene cluster → encodes for enzymes that synthesize secondary metabolite



*Amycolactopsis orientalis*

Source: DSMZ-German Collection of Microorganisms and Cell Cultures GmbH



vancomycin

To reiterate this process: students look up what organism(s) produce that secondary metabolite, find its genomic sequence, use antiSMASH to identify gene clusters

Bacterial Secondary Metabolite	Student Name	Natural Product class	Bacterial Organism	NCBI genome nucleotide accession number(s)
erythromycin				
zwittermicin A				
oxytetracycline				
... etc. and students can add their own				

Genome

Genome

saccharopolyspora erythraea[orgn]

Search

Help

Create alert Limits Advanced

COVID-19 is an emerging, rapidly evolving situation.

Get the latest public health information from CDC: <https://www.coronavirus.gov>.Get the latest research from NIH: <https://www.nih.gov/coronavirus>.**Saccharopolyspora erythraea****Representative genome:** [Saccharopolyspora erythraea NRRL 2338](#)Download sequences in FASTA format for [genome](#), [protein](#)Download genome annotation in [GFF](#), [GenBank](#) or [tabular](#) formatBLAST against Saccharopolyspora erythraea [genome](#), [protein](#)**All 4 genomes for species:**Browse the [list](#)Download sequence and annotation from [RefSeq](#) or [GenBank](#)**Tools**[BLAST Genome](#)**Related information**[Assembly](#)[BioProject](#)[Components](#)[Protein](#)[PubMed](#)[Taxonomy](#)**Search details****"Saccharopolyspora erythraea"**  
[Organism]

Search

See more...

**Recent activity**[Turn Off](#) [Clear](#)[Saccharopolyspora erythraea](#)

Genome

[saccharopolyspora erythraea\[orgn\] \(1\)](#)

Genome

[Streptomyces rimosus subsp. rimosus](#)

Genome

[Streptomyces rimosus](#)

Genome



Display Settings: Overview

Send to:

ID: 688

[Organism Overview](#) ; [Genome Assembly and Annotation report \[4\]](#)

## Saccharopolyspora erythraea

Source of the antibiotic erythromycin

Lineage: [Bacteria\[26684\]](#); [Actinobacteria\[3000\]](#); [Actinobacteria\[2696\]](#); [Pseudonocardiales\[179\]](#); [Pseudonocardiaceae\[178\]](#); [Saccharopolyspora\[11\]](#); [Saccharopolyspora erythraea\[1\]](#)**Saccharopolyspora erythraea.** *Saccharopolyspora erythraea* is the soil bacterium that produces the industrially important antibiotic erythromycin. A. Erythromycin is a clinically important and potent macrolide antibiotic. It is used to treat infections caused by several prokaryotic pathogens such as *Streptococcus*, *Staphylococcus*. [More...](#)

### Summary

**Sequence data:** genome assemblies: 4 (See [Genome Assembly and Annotation report](#))**Statistics:** median total length (Mb): 8.14594  
median protein count: 6901  
median GC%: 71.15

### Publications (limited to 20 most recent records)

1. [Genome Sequence of \*Saccharopolyspora erythraea\* D, a Hyperproducer of Erythromycin](#). Liu WB, et al. *Genome Announc* 2013 Sep 19
  2. [Complete genome sequence of the erythromycin-producing bacterium \*Saccharopolyspora erythraea\* NRRL2338](#). Oliynyk M, et al. *Nat Biotechnol* 2007 Apr
  3. [The erythromycin biosynthetic gene cluster of \*Aeromicrobium erythreum\*](#). Brikun IA, et al. *J Ind Microbiol Biotechnol* 2004 Aug
- [More...](#)

### Representative (genome information for reference and representative genomes)

want nucleotide “RefSeq” or accession,  
usually starts with NC\_ ##### or NZ\_ #####

only use for “Level” complete



NCBI Resources How To Sign in to NCBI

Genome Genome Search Limits Advanced Help

COVID-19 is an emerging, rapidly evolving situation.  
Get the latest public health information from CDC: <https://www.coronavirus.gov>.  
Get the latest research from NIH: <https://www.nih.gov/coronavirus>.

Organism Overview ; Genome Assembly and Annotation report [4]

### Saccharopolyspora erythraea

Levels:  All  Complete [1]  Contig [3] Download table

Organism/Name	Strain	BioSample	BioProject	Assembly	Level	Size (Mb)	GC%	Replicons	WGS	Scaffolds	Gene	Protein	Release Date	Modify Date	FTP
Saccharopolyspora erythraea NRRL 2338	NRRL 2338	SAMEA2272514	PRJEA18489	GCA_000062885.1	●	8.2128	71.10	chromosome:NC_009142.1/AM420293.1	-	-	7257	6976	2007/03/21	2020/02/14	◆◆
Saccharopolyspora erythraea NRRL 2338	DSM 40517	SAMN05192572	PRJNA323074	GCA_002564065.1	○	8.2301	71.10	-	PDBV01	1	7309	6976	2017/10/17	2019/07/30	◆◆
Saccharopolyspora erythraea NRRL 2338	NRRL 2338	SAMN02470651	PRJNA17885	GCA_000171635.1	○	8.07908	71.20	-	ABFV01	241	7243	6826	2009/09/04	2019/12/19	◆◆
Saccharopolyspora erythraea D	D	SAMN02299070	PRJNA214030	GCA_000448385.1	○	7.96992	71.20	-	AVCN01	384	7290	6638	2013/08/20	2019/10/29	◆◆

Items 1 - 4 of 4 << First < Prev Page 1 of 1 Next > Last >>

You are here: NCBI > Genomes & Maps > Genome Support Center

GETTING STARTED RESOURCES POPULAR FEATURED NCBI INFORMATION

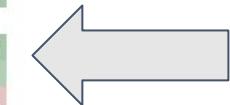
# erythromycin from *Saccharopolyspora erythraea* (NC\_009142.1)



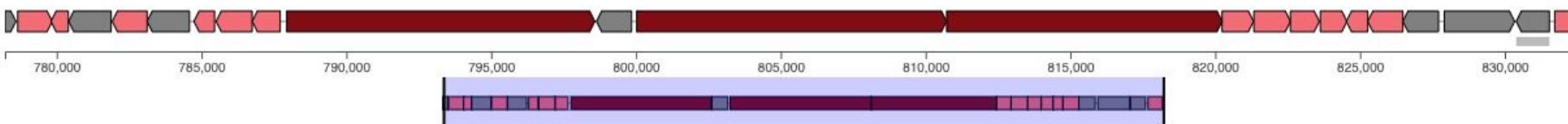
## Identified secondary metabolite regions using strictness 'relaxed'

NC\_009142.1 (Saccharopolyspora erythraea NRRL 2338)

Region	Type	From	To	Most similar known cluster	Polyketide	Similarity
Region 1	hglE-KS , T1PKS	10,554	61,359	nataxazole	Polyketide	7%
Region 2	ectoine	530,970	541,377	ectoine	Other	100%
Region 3	arylpolyene	651,603	692,206			
Region 4	T1PKS	769,006	839,270	erythromycin A / erythromycin B / erythromycin C / erythromycin D	Polyketide:Modular type I + Saccharide:Hybrid/tailoring	100%
Region 5	T3PKS	1,346,130	1,386,323	flaviolin rhamnoside / 3,3'-diflaviolin / flaviolin	Polyketide:Type III + Saccharide:Oligosaccharide	88%
Region 6	NRPS	1,408,067	1,461,769	marformycin A / marformycin B / marformycin C / marformycin D / marformycin E / marformycin F	NRPs	16%
Region 7	indole	2,128,535	2,149,623	fortimicin	Saccharide	6%
Region 8	terpene	2,365,339	2,383,720	SF2575	Polyketide:Type II + Saccharide:Hybrid/tailoring	4%
Region 9	T1PKS	2,511,613	2,561,344	FR-900520	NRp + Polyketide	18%
Region 10	T1PKS	2,781,996	2,821,963	esmeraldin	Polyketide + Other:Aminocoumarin	8%
Region 11	NRPS , T1PKS	2,822,338	2,895,287	heronamide A / heronamide B / heronamide C / heronamide D / heronamide E / heronamide F	NRp + Polyketide	20%
Region 12	NRPS	2,926,118	2,980,854	amychelin	NRp	43%
Region 13	T1PKS , NRPS-like	3,122,590	3,170,855	microansamycin	Polyketide	46%
Region 14	arylpolyene	3,232,357	3,272,561			
Region 15	NRPS-like , NRPS , betalactone	3,285,172	3,381,668	erythrochelin	NRp	100%
Region 16	terpene	3,510,285	3,531,525	geosmin	Terpene	100%
Region 17	NRPS-like	3,545,920	3,588,328	actinomycin D	NRp	14%
Region 18	terpene	3,604,373	3,623,268	isorenieratene	Terpene	57%
Region 19	terpene	3,888,060	3,908,621	neoantimycin	NRp + Polyketide	20%
Region 20	terpene	4,066,597	4,086,366	2-methylisoborneol	Terpene	100%
Region 21	terpene	4,352,437	4,373,977	geosmin	Terpene	100%
Region 22	lanthipeptide	4,414,550	4,434,596	labyrinthopeptin A2 / labyrinthopeptin A1 / labyrinthopeptin A3	RiPP:Lanthipeptide	40%
Region 23	T1PKS	4,547,182	4,631,288	linfuranone B / linfuranone C	Polyketide	61%
Region 24	NRPS , T1PKS , terpene	4,740,186	4,863,811	hopene	Terpene	46%
Region 25	lanthipeptide	4,908,211	4,928,961			
Region 26	hglE-KS	4,977,824	5,020,592	oxalomycin B	NRp + Polyketide	6%
Region 27	T3PKS , PKS-like	5,089,056	5,135,049	meilingmycin	Polyketide	10%
Region 28	lanthipeptide	5,148,567	5,171,161	yatakemycin	NRp	6%
Region 29	terpene	5,188,354	5,207,640	brasiliocardin A	Terpene + Saccharide	38%
Region 30	terpene , NRPS-like	5,471,054	5,519,546	JBIR-126	NRp	18%
Region 31	thiopeptide	5,526,147	5,564,056	azinomycin B	NRp + Polyketide	6%
Region 32	PKS-like , T1PKS	5,912,907	5,958,471	polyketomycin	Polyketide:Iterative type I + Polyketide:Type II + Saccharide:Hybrid/tailoring	4%
Region 33	T1PKS	6,205,358	6,250,574	branched-chain fatty acids	Other	75%
Region 34	other	6,271,927	6,312,481			
Region 35	arylpolyene	6,653,329	6,694,477			
Region 36	terpene	7,101,625	7,122,512			



Location: 778,214 - 832,825 nt. (total: 54,612 nt).  
 This entry is originally from NCBI GenBank AM420293.1.

[Download Cluster GenBank file](#)[View antiSMASH-generated output](#)

Select a gene to view the details available for it

## Legend:

[reset view](#)[zoom to selection](#)
[General](#) [Compounds](#) [Genes](#) [Polyketide](#) [Saccharide](#) [History](#) [KnownClusterBlast](#) [NRPS/PKS domains](#)

General information about the BGC

MIBiG accession

BGC0000055

Short description

erythromycin A biosynthetic gene cluster from *Saccharopolyspora erythraea* NRRL 2338

Status

Minimal annotation: no [?](#)Completeness: complete [?](#)

Biosynthetic class(es)

Polyketide (Macrolide)

Loci

Saccharide (hybrid/tailoring)

NCBI GenBank: [AM420293.1](#)

Compounds

erythromycin A

erythromycin B

erythromycin C

erythromycin D

Species

*Saccharopolyspora erythraea* NRRL 2338 [taxonony]

- [Type I polyketide biosynthesis in bacteria \(Part A-erythromycin biosynthesis\).](#) Rawlings BJ et al., Nat Prod Rep (2001) PMID:11336289
- [Architectures of whole-module and bimodular proteins from the 6-deoxyerythronolide B synthase.](#) Edwards AL et al., J. Mol. Biol. (2014) PMID:24704088
- [Systems perspectives on erythromycin biosynthesis by comparative genomic and transcriptomic analyses of \*S. erythraea\* E3 and NRRL2338 strains.](#) Li YY et al., BMC Genomics (2013) PMID:23902230
- [Knocking out of tailoring genes eryK and eryG in an industrial erythromycin-producing strain of \*Saccharopolyspora erythraea\* leading to overproduction of erythromycin B, C and D at different conversion ratios.](#) Zhang Q et al., Lett. Appl. Microbiol. (2011) PMID:21175699
- [Knockout of the erythromycin biosynthetic cluster gene, eryB1, blocks isoflavone glucoside bioconversion during erythromycin fermentations in \*Aeromicrobium erythreum\* but not in \*Saccharopolyspora erythraea\*.](#) Reeves AR et al., Appl. Environ. Microbiol. (2008) PMID:18836015
- [Genetic modulation of the overexpression of tailoring genes eryK and eryG leading to the improvement of erythromycin A purity and production in \*Saccharopolyspora erythraea\* fermentation.](#) Chen Y et al., Appl. Environ. Microbiol. (2008) PMID:18223111
- [Complete genome sequence of the erythromycin-producing bacterium \*Saccharopolyspora erythraea\* NRRL2338.](#) Oliynyk M et al., Nat. Biotechnol. (2007) PMID:17369815
- [Effects of methylmalonyl-CoA mutase gene knockouts on erythromycin production in carbohydrate-based and oil-based fermentations of \*Saccharopolyspora erythraea\*.](#) Reeves AR et al., J. Ind. Microbiol. Biotechnol. (2006) PMID:16491356
- [Improved bioconversion of 15-fluoro-6-deoxyerythronolide B to 15-fluoro-erythromycin A by overexpression of the eryK Gene in \*Saccharopolyspora erythraea\*.](#) Desai RP et al., Biotechnol. Prog. (2004 Nov-Dec) PMID:15575696
- [Reconstitution and characterization of a new desosaminyl transferase, EryCII, from the erythromycin biosynthetic pathway.](#) Lee HY et al., J. Am. Chem. Soc. (2004) PMID:15303858
- [A specific role of the \*Saccharopolyspora erythraea\* thioesterase II gene in the function of modular polyketide synthases.](#) Hu Z et al., Microbiology (Reading, Engl.) (2003) PMID:12904561
- [Analysis of an 8.1-kb DNA fragment contiguous with the erythromycin gene cluster of \*Saccharopolyspora erythraea\* in the eryCl-flanking region.](#) Reeves AR et al., Antimicrob. Agents Chemother. (2002) PMID:12435693
- [Transcriptional organization of the erythromycin biosynthetic gene cluster of \*Saccharopolyspora erythraea\*.](#) Reeves AR et al., J. Bacteriol. (1999) PMID:10559177
- [Analysis of eryB1, eryBIII and eryBVII from the erythromycin biosynthetic gene cluster in \*Saccharopolyspora erythraea\*.](#) Gaisser S et al., Mol. Gen. Genet. (1998) PMID:9613575
- [Targeted gene inactivation for the elucidation of deoxysugar biosynthesis in the erythromycin producer \*Saccharopolyspora erythraea\*.](#) Salah-Bey K et al., Mol. Gen. Genet. (1998) PMID:9563840
- [The loading domain of the erythromycin polyketide synthase is not essential for erythromycin biosynthesis in \*Saccharopolyspora erythraea\*.](#) Pereda A et al., Microbiology (Reading, Engl.) (1998) PMID:9493390

References

25

# Your turn to use antiSMASH!

1. Marc will screen share and walk you through submitting a job.
  - a. refer to our eHandout!
  - b. he will explain a bit about submission parameters
2. We will use Zoom breakout rooms so you can have a small, interactive peer group while you try using antiSMASH.
3. Use antiSMASH to identify BGCs in bacterial genome of your choice:
  - a. genome from organism of your choice ex. top species BLAST match of one of your TE isolates,
  - b. a genome Marc has shared from TECH,
  - c. genome of organism that makes a well-known secondary metabolite (Table in eHandout),
  - d. genome of organism that makes secondary metabolite of interest to you.
4. Share your results in Google Slides (link in chat).



# Another example: oxytetracycline from *Streptomyces rimosus* (NZ\_CP025551.1)

antiSMASH antiSMASH version 5.1.2 Download About Help Contact

Select genomic region: Overview 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9 1.10 1.11 1.12 1.13 1.14 1.15 1.16 1.17 1.18 1.19 1.20 1.21 1.22 1.23 1.24 1.25 1.26 1.27 1.28 1.29 1.30 1.31 1.32 1.33 1.34 1.35 1.36 1.37 1.38 1.39 1.40 1.41 1.42 1.43 1.44 1.45

Identified secondary metabolite regions using strictness 'relaxed'

NZ\_CP023688.1 (*Streptomyces rimosus*)

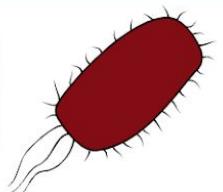
Region	Type	From	To	Most similar known cluster	Similarity
Region 1	NRPS-like	75,861	117,611	paromomycin	Saccharide 7%
Region 2	NRPS, T1PKS, terpene	178,391	268,367	isorenieratene	Terpene 85%
Region 3	T1PKS	308,500	367,400	sceliphrolactam	Polyketide 32%
Region 4	T1PKS, NRPS-like	389,996	509,176	nystatin A1	Polyketide:Modular type I + Saccharide:Hybrid/tailoring 72%
Region 5	NRPS	518,914	562,807	qinichelins	NRP 22%
Region 6	lassopeptide	575,880	596,639	lagmysin	RiPP 80%
Region 7	T2PKS, terpene	598,260	673,898	oxytetracycline	Polyketide 95%
Region 8	T1PKS	769,064	813,615	A54145	NRP 3%
Region 9	lanthipeptide	893,989	915,856		
Region 10	T1PKS	916,353	961,296	spiroindimicin A / spiroindimicin B / spiroindimicin C / spiroindimicin D / indimicin A / indimicin B / indimicin C / indimicin D / indimicin E / lynamycin A / lynamycin D / lynamycin F / lynamycin G	Other 6%
Region 11	NRPS	988,923	1,079,016	rimosamide	NRP 92%

Region	Molecule Type	Count	Count	Molecule	Molecule Type	Percentage
Region 11	NRPS ↗	988,923	1,079,016	rimosamide ↗	NRP	92%
Region 12	NRPS ↗	1,096,942	1,142,296	daptomycin ↗	NRP	14%
Region 13	arylpolyene ↗, terpene ↗	1,164,799	1,219,268	herboxidiene ↗	Polyketide	3%
Region 14	terpene ↗	1,384,859	1,410,805	hopene ↗	Terpene	76%
Region 15	NRPS ↗	1,562,013	1,637,059	isocomplestatin ↗	NRP	93%
Region 16	terpene ↗	1,699,324	1,720,042	kanamycin ↗	Saccharide	5%
Region 17	melanin ↗	1,756,440	1,767,825			
Region 18	other ↗	1,819,199	1,860,249	A-503083 A / A-503083 B / A-503083 E / A-503083 F ↗	NRP	7%
Region 19	oligosaccharide ↗	2,008,736	2,041,015			
Region 20	bacteriocin ↗	2,073,737	2,085,092			
Region 21	butyrolactone ↗	2,160,114	2,169,944			
Region 22	lanthipeptide ↗	2,185,371	2,207,571			
Region 23	nucleoside ↗	2,210,916	2,231,039	tubercidin ↗	Other	27%
Region 24	NRPS ↗, indole ↗, betalactone ↗	2,250,600	2,413,822	streptobactin ↗	NRP	70%
Region 25	siderophore ↗	2,502,370	2,515,871	ficellomycin ↗	NRP	3%
Region 26	NRPS ↗, PKS-like ↗	3,074,948	3,131,267	tyrobetaine ↗	NRP	100%
Region 27	NRPS ↗	4,130,265	4,211,281	mannopeptimycin ↗	NRP	22%
Region 28	arylpolyene ↗	4,248,476	4,289,636	fusaricidin B ↗	Polyketide + NRP:Lipopeptide	25%
Region 29	NRPS ↗	4,783,066	4,835,578	ishigamide ↗	NRP + Polyketide	61%
Region 30	butyrolactone ↗	5,452,364	5,463,088	neocarzinostatin ↗	Polyketide:Iterative type I + Polyketide:Enediyne type I	8%
Region 31	lassopeptide ↗	5,823,027	5,845,574	moomysin ↗	RiPP	50%
Region 32	lanthipeptide ↗	6,575,665	6,598,301	SAL-2242 ↗	RiPP:Lanthipeptide	77%
Region 33	terpene ↗	6,802,593	6,824,290	geosmin ↗	Terpene	100%
Region 34	ectoine ↗	7,235,577	7,245,999	ectoine ↗	Other	100%
Region 35	siderophore ↗	7,323,848	7,331,429	desferrioxamine E ↗	Other	39 100%

Region	Metabolite Type	Count	Count	Metabolite	Metabolite Type	Percentage
Region 32	lanthipeptide ↗	6,575,665	6,598,301	SAL-2242 ↗	RiPP:Lanthipeptide	77%
Region 33	terpene ↗	6,802,593	6,824,290	geosmin ↗	Terpene	100%
Region 34	ectoine ↗	7,235,577	7,245,999	ectoine ↗	Other	100%
Region 35	siderophore ↗	7,323,848	7,331,429	desferrioxamine E ↗	Other	100%
Region 36	siderophore ↗	7,427,099	7,440,833			
Region 37	terpene ↗	8,040,869	8,060,614			
Region 38	NRPS ↗ , T1PKS ↗	8,335,930	8,388,892	marinacaroline A / marinacaroline B / marinacaroline C / marinacaroline D ↗	Alkaloid	23%
Region 39	NRPS-like ↗	8,396,187	8,440,103			
Region 40	NRPS ↗ , phosphonate ↗	8,480,169	8,548,216	deimino-antipain ↗	NRP	66%
Region 41	NRPS ↗ , NRPS-like ↗ , T1PKS ↗ , terpene ↗	8,600,669	8,727,300	tetronasin ↗	Polyketide	9%
Region 42	other ↗ , NRPS-like ↗	8,804,014	8,875,702	A83543A ↗	Polyketide	8%
Region 43	butyrolactone ↗	8,881,642	8,892,628	cyphomycin ↗	Polyketide	11%
Region 44	T1PKS ↗ , NRPS ↗ , other ↗ , nucleoside ↗	8,964,456	9,150,779	pseudouridimycin ↗	Other:Nucleoside	68%
Region 45	NRPS ↗	9,233,008	9,282,905			

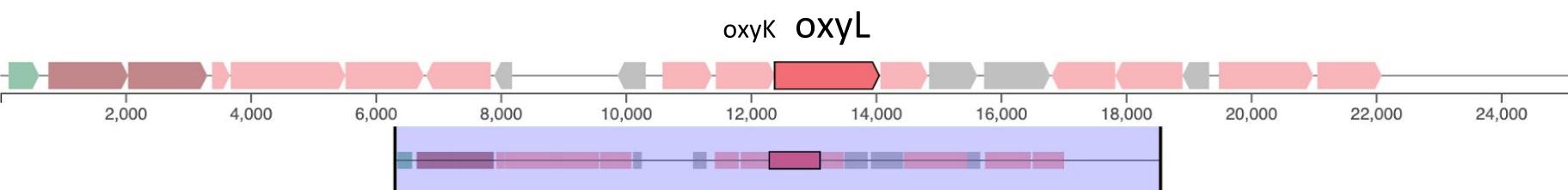
anti  
SMASH

If you have found antiSMASH useful, please cite us .



**BGC0000254: oxytetracycline biosynthetic gene cluster from *Streptomyces rimosus*****Gene details**

Location: 1 - 25,222 nt. (total: 25,222 nt).

This entry is originally from NCBI GenBank [DQ143963.2](#).[Download Cluster GenBank file](#)[View antiSMASH-generated output](#)**Legend:**

- core biosynthetic genes
  - additional biosynthetic genes
  - transport-related genes
  - regulatory genes
  - resistance genes
  - other genes
- [reset view](#)   [zoom to selection](#)

**oxyL**

Locus tag: None

Protein ID: AAZ78335.1

Gene: oxyL

Location: 12,386 - 14,059, (total: 1674 nt)

**Functions:**

-

biosynthetic-additional (smcogs)  
 SMCOG1050:monooxygenase FAD-binding  
 (Score: 380.2; E-value: 2.9e-115)  
[NCBI BlastP on this gene](#)  
[View genomic context](#)  
[MiBIG Hits](#)  
 AA sequence: [Copy to clipboard](#)  
 Nucleotide sequence: [Copy to clipboard](#)

[General](#) [Compounds](#) [Genes](#) [History](#) [KnownClusterBlast](#) [NRPS/PKS domains](#)

General information about the BGC

MiBIG accession

BGC0000254

Short description

oxytetracycline biosynthetic gene cluster from *Streptomyces rimosus*

Status

Minimal annotation: yes [?](#)

Biosynthetic class(es)

Polyketide

Loci

NCBI GenBank: [DQ143963.2](#)

Compounds

[oxytetracycline](#)

Species

*Streptomyces rimosus* [taxonomy]

References

[Engineered biosynthesis of a novel amidated polyketide, using the malonamyl-specific initiation module from the oxytetracycline polyketide synthase.](#) Zhang W et al., Appl. Environ. Microbiol. (2006) PMID:16597959

Nucleotide Nucleotide Advanced Help

COVID-19 is an emerging, rapidly evolving situation.  
Get the latest public health information from CDC: <https://www.coronavirus.gov>.  
Get the latest research from NIH: <https://www.nih.gov/coronavirus>.

GenBank (full) ▾

### Streptomyces coelicolor A3(2) chromosome, complete genome

NCBI Reference Sequence: NC\_003888.3

[FASTA](#) [Graphics](#)

Go to:

LOCUS	NC_003888	8667507 bp	DNA	linear	CON 03-AUG-2016
DEFINITION	Streptomyces coelicolor A3(2) chromosome, complete genome.				
ACCESSION	NC_003888				
VERSION	NC_003888.3				
DBLINK	BioProject: <a href="#">PRJNA57801</a> Assembly: <a href="#">GCF_000203835.1</a>				
KEYWORDS	RefSeq; complete genome.				
SOURCE	Streptomyces coelicolor A3(2)				
ORGANISM	Streptomyces coelicolor A3(2) Bacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; Streptomyces; Streptomyces albidoflavus group.				
REFERENCE	1 (bases 1 to 8667507)				
AUTHORS	Hsiao,N.H. and Kirby,R.				
TITLE	Comparative genomics of Streptomyces avermitilis, Streptomyces cattleya, Streptomyces maritimus and Kitasatospora aureofaciens using a Streptomyces coelicolor microarray system				
JOURNAL	Antonie Van Leeuwenhoek 93 (1-2), 1-25 (2008)				
PUBMED	<a href="#">17588127</a>				
REFERENCE	2 (bases 1 to 8667507)				
AUTHORS	Bentley,S.D., Chater,K.F., Cerdeno-Tarrega,A.M., Challis,G.L., Thomson,N.R., James,K.D., Harris,D.E., Quail,M.A., Kieser,H., Harper,D., Bateman,A., Brown,S., Chandra,G., Chen,C.W., Collins,M., Cronin,A., Fraser,A., Goble,A., Hidalgo,J., Hornsby,T., Howarth,S., Huang,C.H., Kieser,T., Larke,L., Murphy,L., Oliver,K., O'Neil,S., Rabinowitsch,E., Rajandream,M.A., Rutherford,K., Rutter,S., Seeger,K., Saunders,D., Sharp,S., Squares,R., Squares,S., Taylor,K., Warren,T., Wietzorek,A., Woodward,J., Barrell,B.G., Parkhill,J. and Hopwood,D.A.				
TITLE	Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)				
JOURNAL	Nature 417 (6885), 141-147 (2002)				
PUBMED	<a href="#">12000953</a>				
REFERENCE	3 (bases 1 to 8667507)				
AUTHORS	Redenbach,M., Kieser,H.M., Denapaité,D., Eichner,A., Cullum,J., Kinashi,H. and Hopwood,D.A.				
TITLE	A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome				
JOURNAL	Mol. Microbiol. 21 (1), 77-96 (1996)				
PUBMED	<a href="#">8843436</a>				
REFERENCE	4 (bases 1 to 8667507)				
CONSRTM	NCBI Genome Project				
TITLE	Direct Submission				
JOURNAL	Submitted (28-MAY-2002) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA				

Send to:

Choose Destination

- Complete Record
- Gene Features

Choose Destination

- File
- Clipboard
- Collections
- Analysis Tool

Download 1 item.

Format: [GenBank \(full\)](#)

Show GI

Create File

Run BLAST

Pick Primers

Highlight Sequence Features

Related information

- Assembly
- BioProject
- PubMed
- Taxonomy
- Components (Core)
- Full text in PMC
- Gene
- Genome
- Identical GenBank Sequence
- Reference Genome BioProject
- Representative Genome BioProject

LinkOut to external resources

- reagents [ExactAntigen/Labome]
- reagents [ExactAntigen/Labome]

If you need to download genome file from NCBI.  
“.gb” file works for upload to antiSMASH.