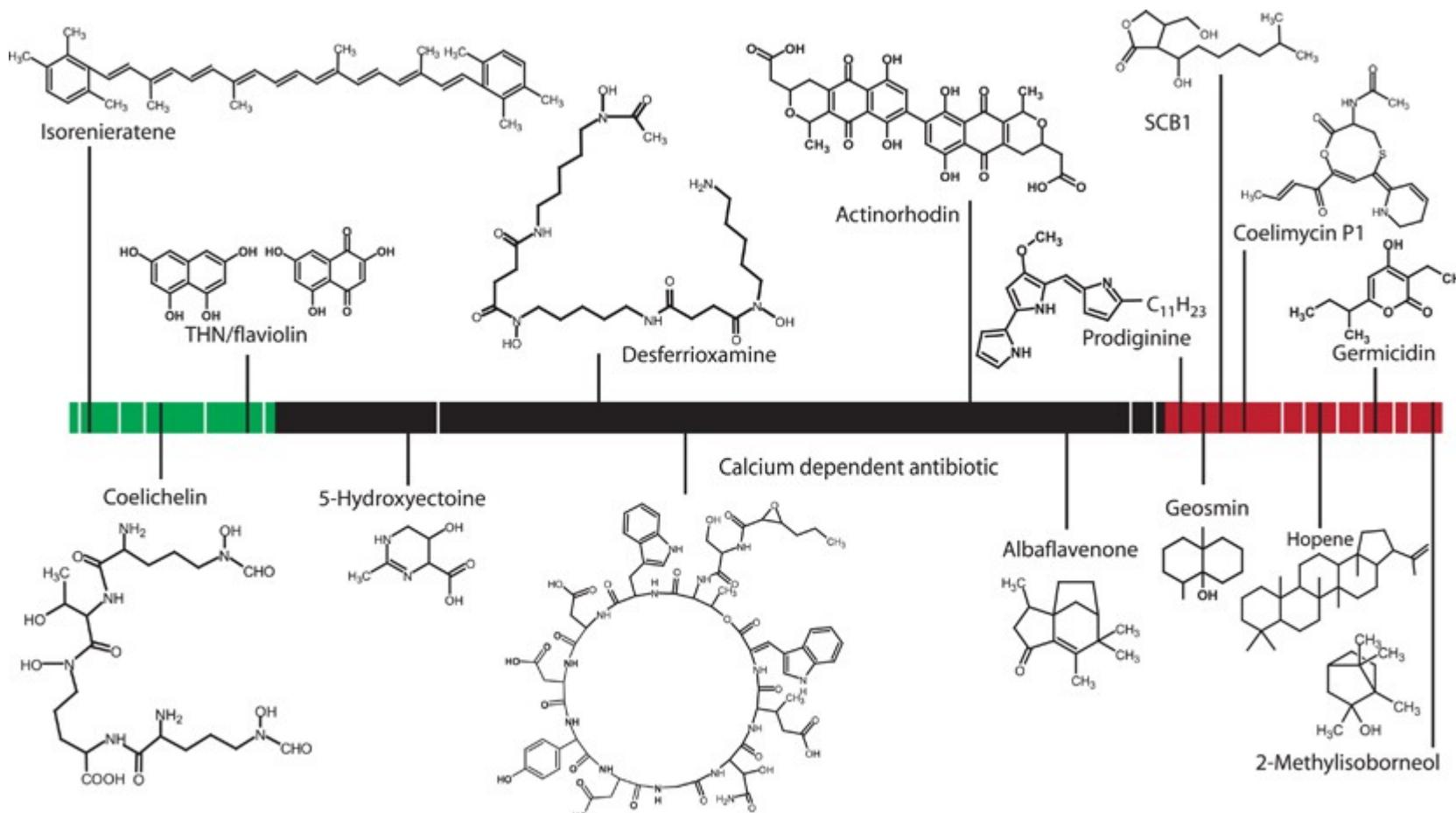


Functional Annotation

Lesson 7: Antibiotics and Secondary Metabolite Analysis
Shell (AntiSMASH)

Instructor: Arkadiy Garber

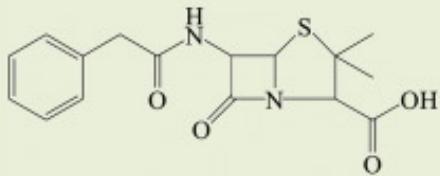
Secondary metabolites: “organic compounds produced by bacteria, fungi, or plants which are not directly involved in the normal growth, development, or reproduction of the organism.” -Wikipedia



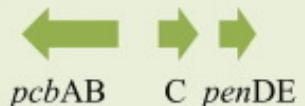
Streptomyces coelicolor secondary metabolites
Arryn Craney, et al., 2013, *Journal of Antibiotics*

Antimicrobials: An antimicrobial is an agent that kills microorganisms or stops their growth (antibiotic is a type of antimicrobial that acts against bacteria)

Penicillin (β -lactam)

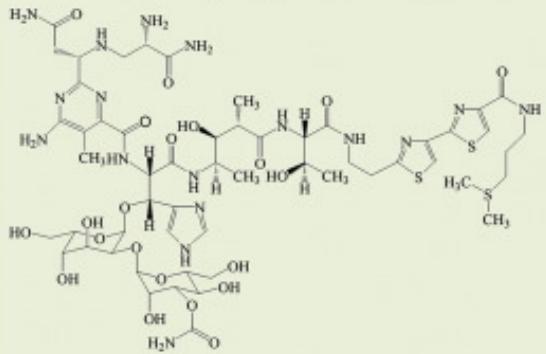


Antibiotic from *Penicillium chrysogenum*



Penicillin biosynthetic gene cluster (17 kb)

Bleomycin (Glycopetide)

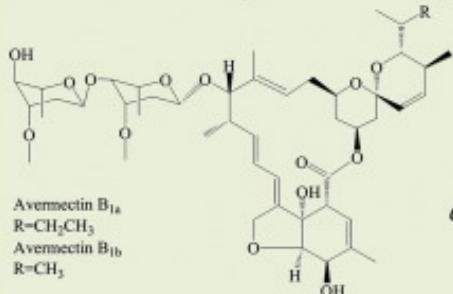


Antitumor agent from *Streptomyces verticillus*

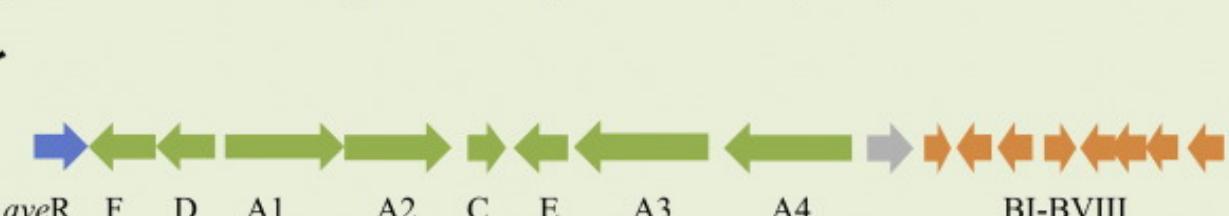


Bleomycin biosynthetic gene cluster (66 kb)

Avermectins (Macrolide)



Antiparasitic agent from *Streptomyces avermitilis*



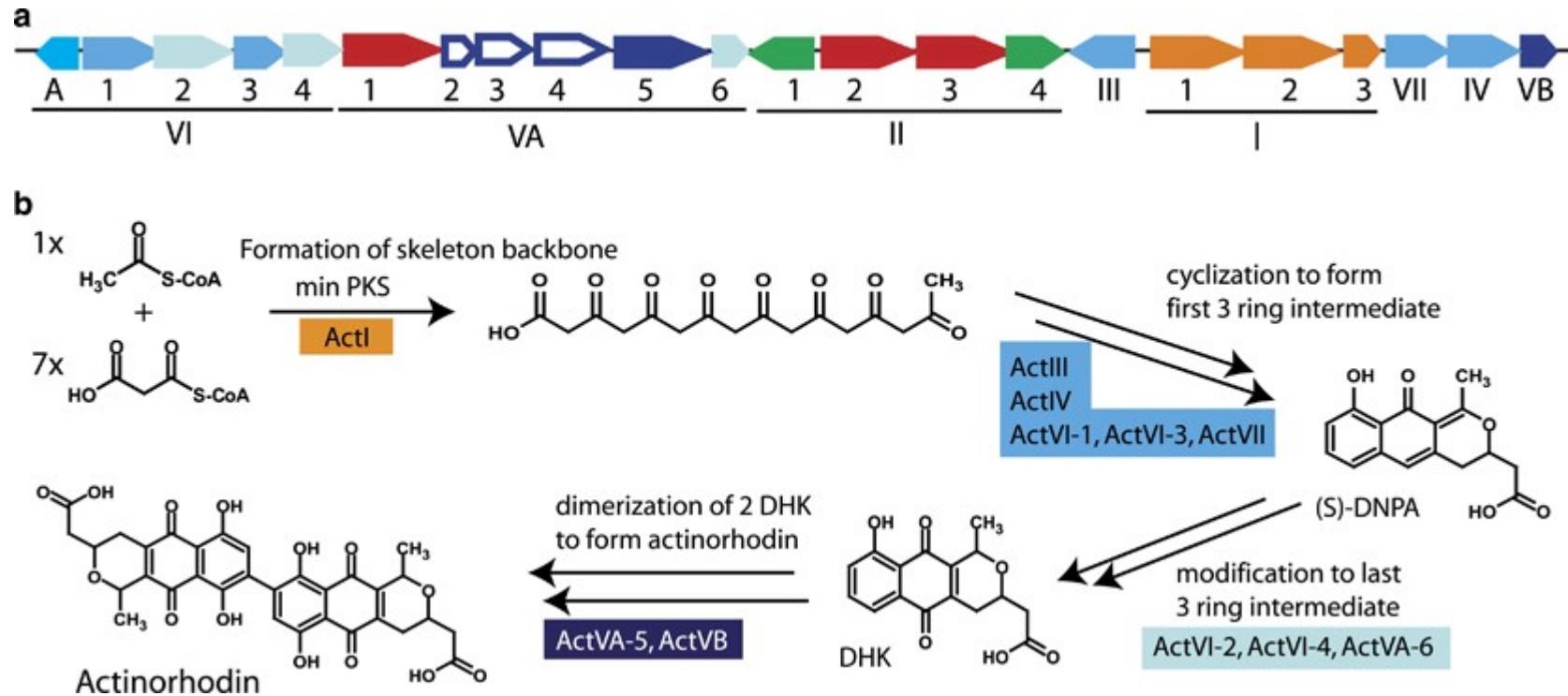
Avermectins biosynthetic gene cluster (82 kb)

Streptomyces coelicolor secondary metabolites

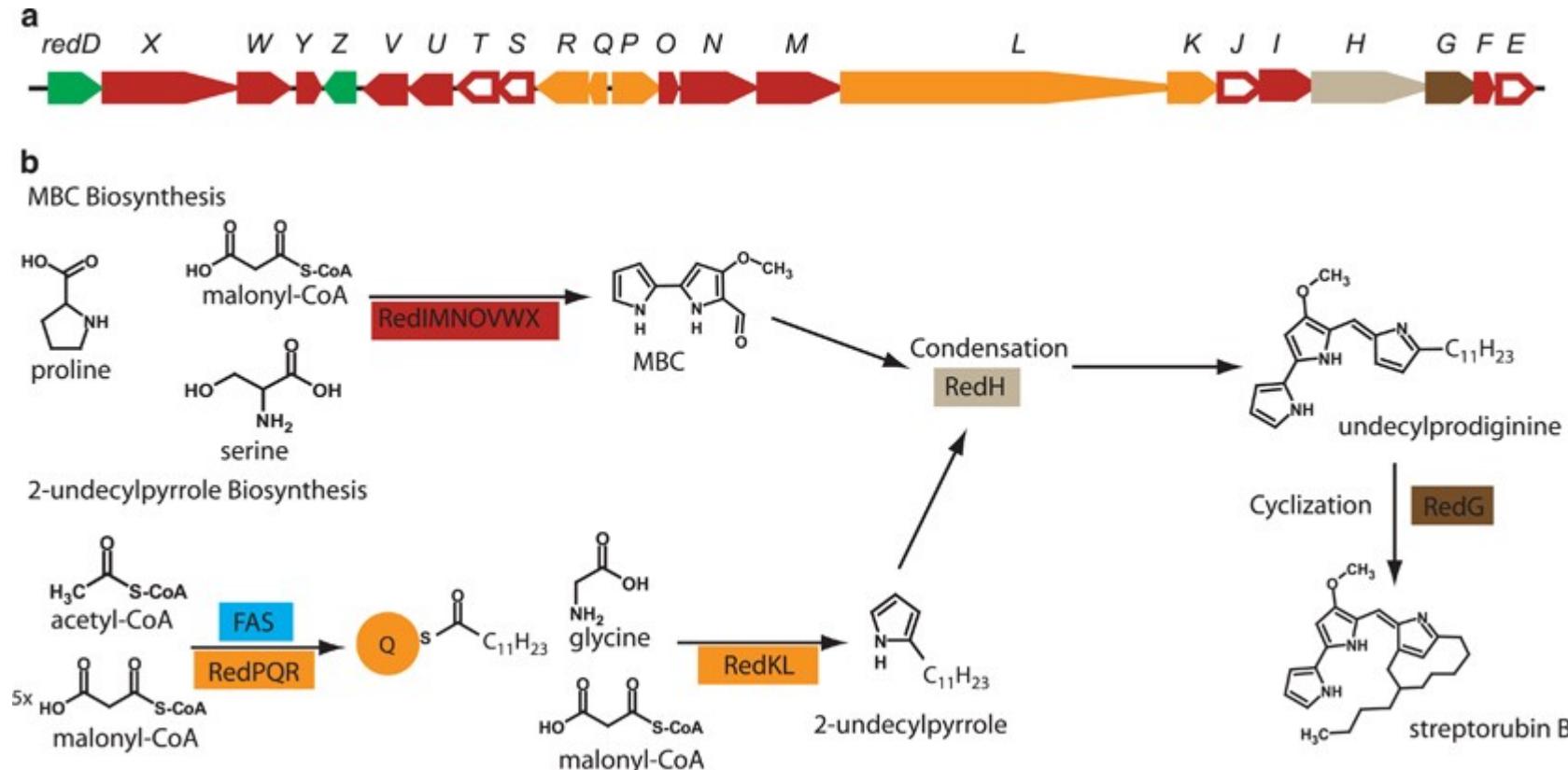
Secondary metabolite	Location	Type	Identification method	Reference
<i>Identified structures</i>				
Isorenieratene	SCO0185-0191	Terpenoid	Blue light induction	63
Coelichelin	SCO0489-0499	NRP	Genome mining	62
THN/flavolin	SCO1206-1208	PK-type III	Genome mining	119
5-Hydroxyectoine	SCO1864-1867	Cyclic amino acid	Salt or high temp	64
Desferrioxamine	SCO2782-2785	Tris-hydroxymate	Genome mining	61
CDA	SCO3210-3249	NRP	Antimicrobial activity	120
Actinorhodin	SCO5071-5092	PK-type II	Blue pigment	20
Albaflavenone	SCO5222-5223	Terpenoid	Odor/genome mining	121
Prodiginine	SCO5877-5898	Tryptrole	Red pigment	21
Geosmin	SCO6073	Terpenoid	Odor/genome mining	122
SCB1	SCO6266	γ -Butyrolactone	Genome mining	123
Coelimycin P1	SCO6273-6288	PK-type I	Yellow pigment/genome mining	36, 124, 125
Hopene	SCO6759-6771	PK-type III	Genome mining	126
Germicidin	SCO7221	PK-type III	Genome mining	127
2-Methylisoborneol	SCO7700-7701	Terpenoid	Odor/genome mining	128
Methylenomycin	SCP1.228c-246	Cyclopentanoid	Antimicrobial activity	129
Methylfurans	SCP1.228c-246	Methylfurans	Genome mining	130
<i>Developmental secondary metabolites</i>				
SapB	SCO6681-6685	Lantibiotic	MS/phenotype	131
<i>Predicted structures (untested)</i>				
Eicosapentaenoic acid	SCO0124-0129	Fatty acid	Genome prediction	13
Melanin	SCO2700-2701	Melanin	Genome prediction	13
Bacteriocin	SCO0753-0756	Bacteriocin	Genome prediction	13
Coelibactin	SCO7681-7691	NRP	Genome prediction	22
<i>Unable to predict structures</i>				
Lantibiotic	SCO0267-0270	Lantibiotic	Genome prediction	13
Lantibiotic	SCO6927-6932	Lantibiotic	Genome prediction	13
PKS	SCO1265-1273	PK-type II	Genome prediction	13
PKS	SCO6826-6827	PK-type II	Genome prediction	13
PKS	SCO7669-7671	PK-type III	Genome prediction	13
Siderophore	SCO5799-5801	—	Genome prediction	13
Dipeptide	SCO6429-6438	—	Genome prediction	13
Gray spore pigment	SCO5314-5320	PK-type II	Gray pigment	132

Abbreviations: NRP, nonribosomal peptide; PK, polyketide.

Actinorhodin biosynthesis: Regulatory genes, resistance genes, minimal PKS (ActI)



Prodiginine biosynthetic cluster



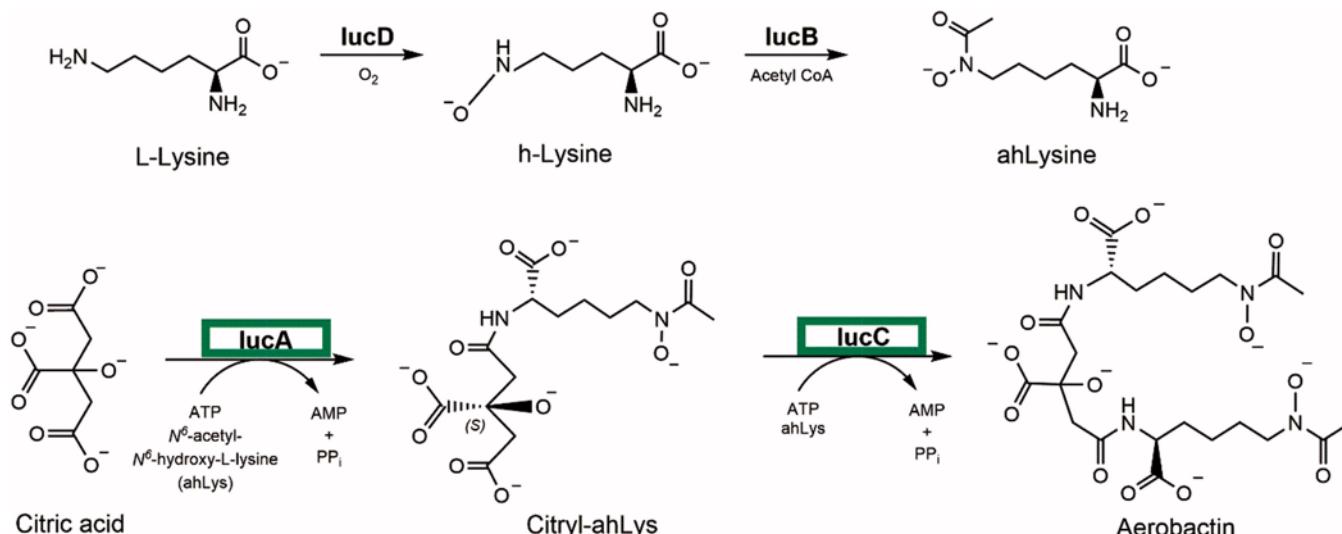
Siderophores - synthesis

What are siderophores?

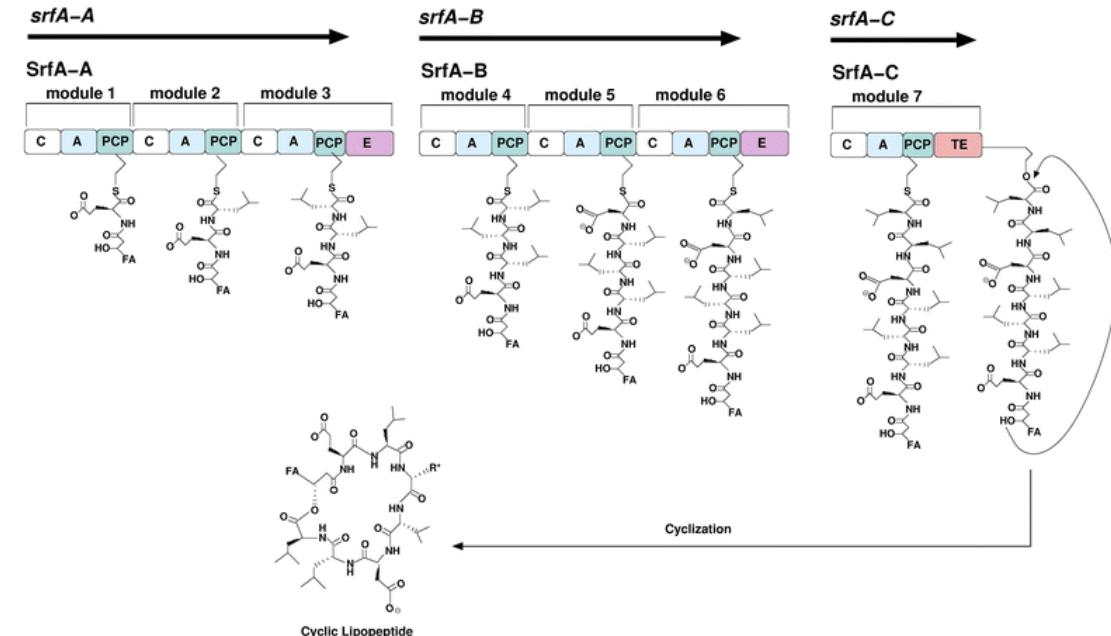
Molecules that are secreted by bacteria and archaea

Bind iron, and the iron(III)-chelated siderophores are then taken up by the cell (sometimes not the same cell that secreted the siderophore!)

NRPS siderophore production

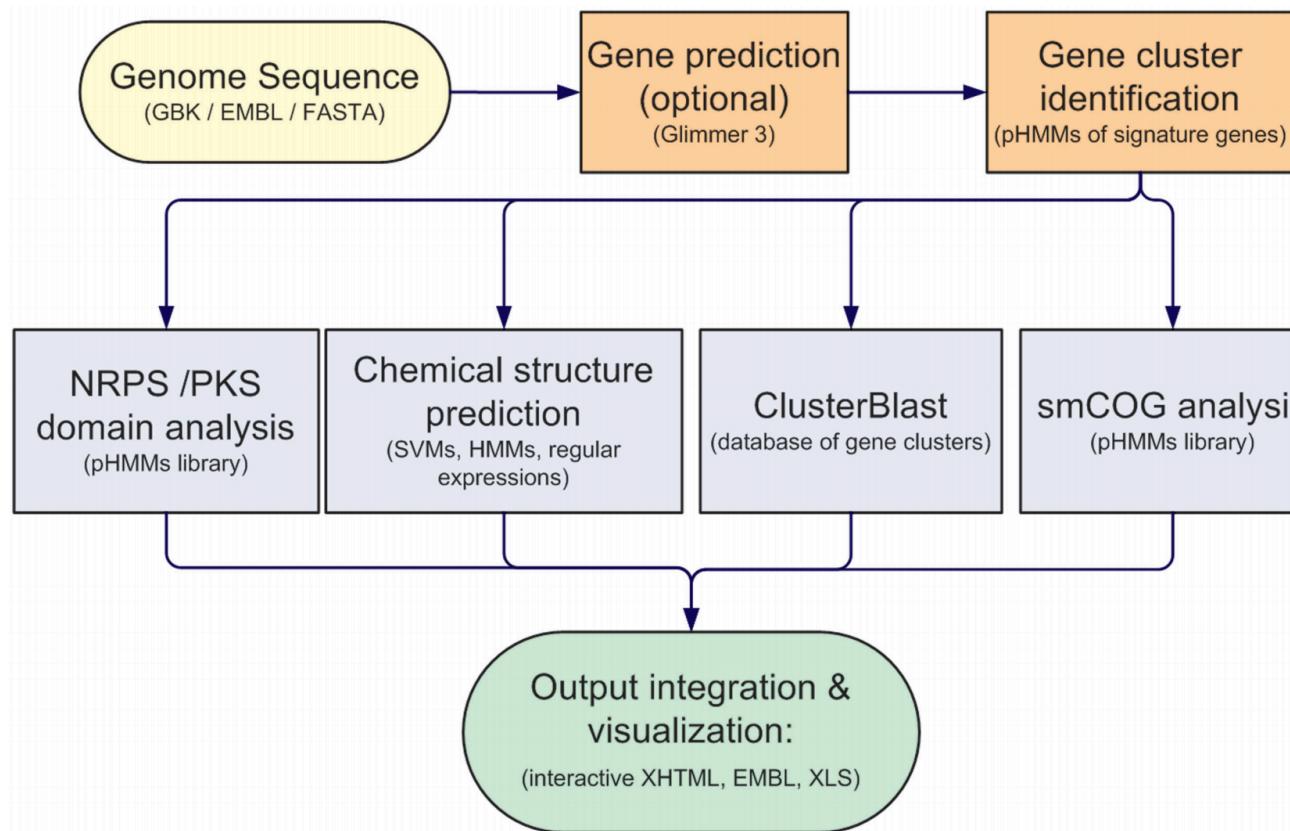


NRPS-independent siderophore production



antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences

Marnix H. Medema^{1,2}, Kai Blin³, Peter Cimermancic⁴, Victor de Jager^{5,6,7},
Piotr Zakrzewski^{1,2}, Michael A. Fischbach⁴, Tilmann Weber³, Eriko Takano^{1,*} and
Rainer Breitling^{2,8}



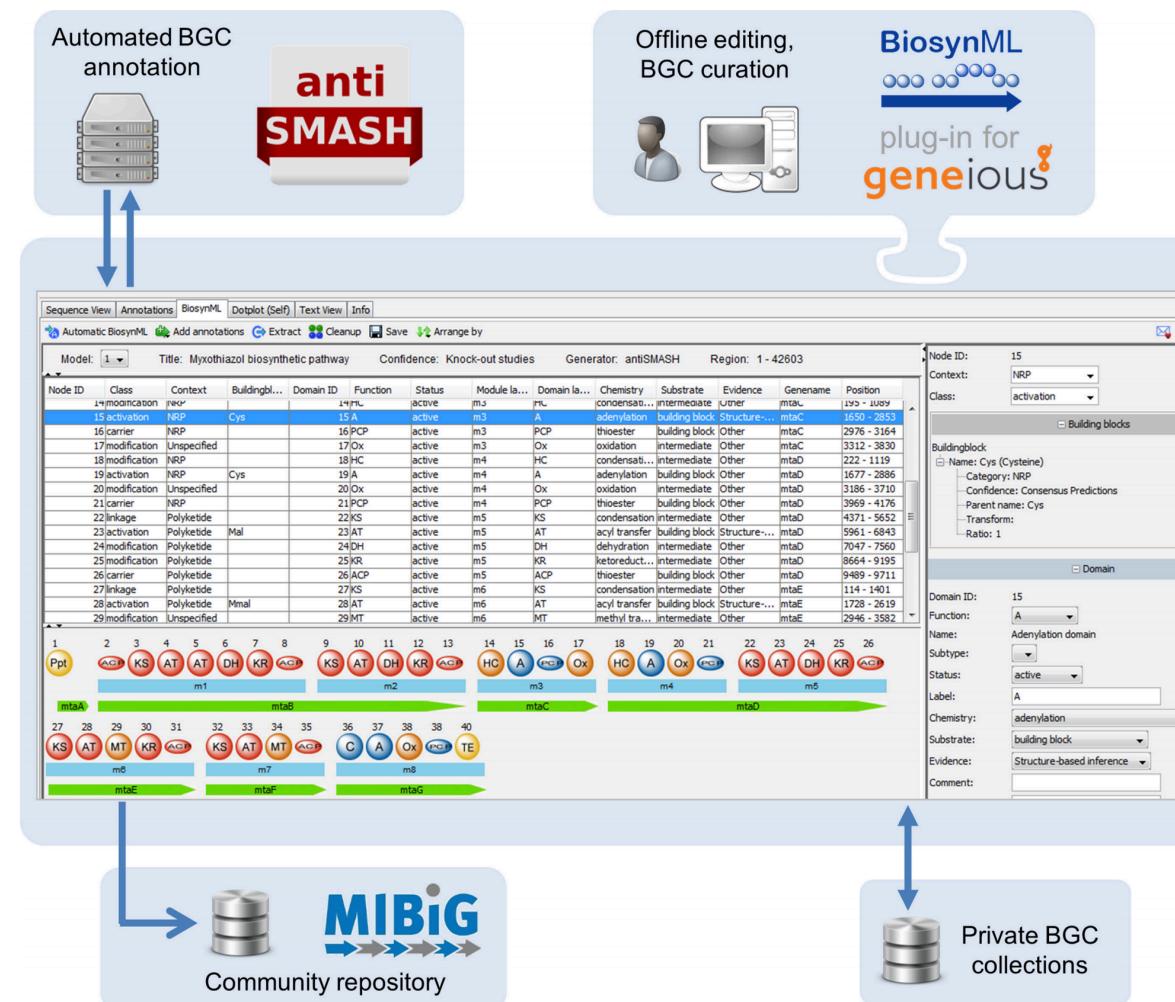
antiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers

Kai Blin¹, Marnix H. Medema^{2,3}, Daniyal Kazempour¹, Michael A. Fischbach⁴,
Rainer Breitling^{3,5,*}, Eriko Takano^{2,5,*} and Tilmann Weber^{1,*}

Features	antiSMASH 2.0	antiSMASH 1.0
Open-source and stand-alone available	X	X
Covers bacteria, archaea and fungi	X	X
NRPS/PKS detection	X	X
NRPS/PKS detailed functional domain annotation	X	X
NRP/PK core structure prediction	X	X
Lantipeptide core structure prediction	X	
Detection of other biosynthetic classes	X	X
Gene cluster border prediction	X	X
Comparative gene cluster analysis	X	X
Sub-cluster analysis	X	
Prediction of putative novel gene cluster types	X	X
Protein sequence input	X	
Nucleotide sequence input	X	X
Multi-contig input	X	
PKS structural modeling		
NRPS/PKS domain phylogenomic analysis	(X) ^a	

antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters

Tilmann Weber^{1,*}, Kai Blin¹, Srikanth Duddela², Daniel Krug^{2,3}, Hyun Uk Kim^{1,4}, Robert Brucolieri⁵, Sang Yup Lee^{1,4}, Michael A. Fischbach⁶, Rolf Müller^{2,3}, Wolfgang Wohlleben^{7,8}, Rainer Breitling⁹, Eriko Takano⁹ and Marnix H. Medema^{10,11,*}



antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification

Kai Blin¹, Thomas Wolf², Marc G. Chevrette³, Xiaowen Lu⁴, Christopher J. Schwalen⁵,
Satria A. Kautsar⁴, Hernando G. Suarez Duran⁴, Emmanuel L. C. de los Santos⁶, Hyun
Uk Kim^{1,7}, Mariana Nave⁸, Jeroen S. Dickschat⁹, Douglas A. Mitchell^{5,10}, Ekaterina Shelest²,
Rainer Breitling¹¹, Eriko Takano¹¹, Sang Yup Lee^{1,7}, Tilmann Weber^{1,*} and Marnix
H. Medema^{4,*}

Rule-based detection of BGCs	Cluster specific analyses
Aminocoumarins	Microcin
Aminoglycosides / aminocyclitols	Microviridin
Aryl polyenes	Non-ribosomal peptides
Bacteriocins	Nucleosides
Beta-lactams	Oligosaccharide
Bottromycin	Others
Butyrolactones	Phenazine
ClusterFinder fatty acid	Phosphoglycolipids
ClusterFinder Saccharide	Phosphonate
Cyanobactins	Polyunsaturated fatty acids
(Dialkyl)resorcinols	Trans-AT type I PKS
Ectoines	Type I PKS
Furan	Type II PKS
<i>Fused (Pheganomycin-like)</i>	Type III PKS
Glycycin	Other (unusual) PKS
Head-to-tail cyclised peptide	Proteusin
Homoserine lactone	Sactipeptide
Indoles	Siderophores
Ladderane lipids	<i>Terpene</i>
Lantipeptides	<i>Thiopeptide</i>
Linear azol(in)e-containing peptides (LAPs)	
<i>Lasso peptide</i>	
Linaridin	
Melanins	
Rule-independent detection of BGCs	
ClusterFinder	<i>ClusterBlast (identification of similar clusters in sequence genomes)</i>
	<i>KnownClusterBlast (identification of similar characterized gene clusters)</i>
	<i>TransAT-PKS Domain Alignments</i>

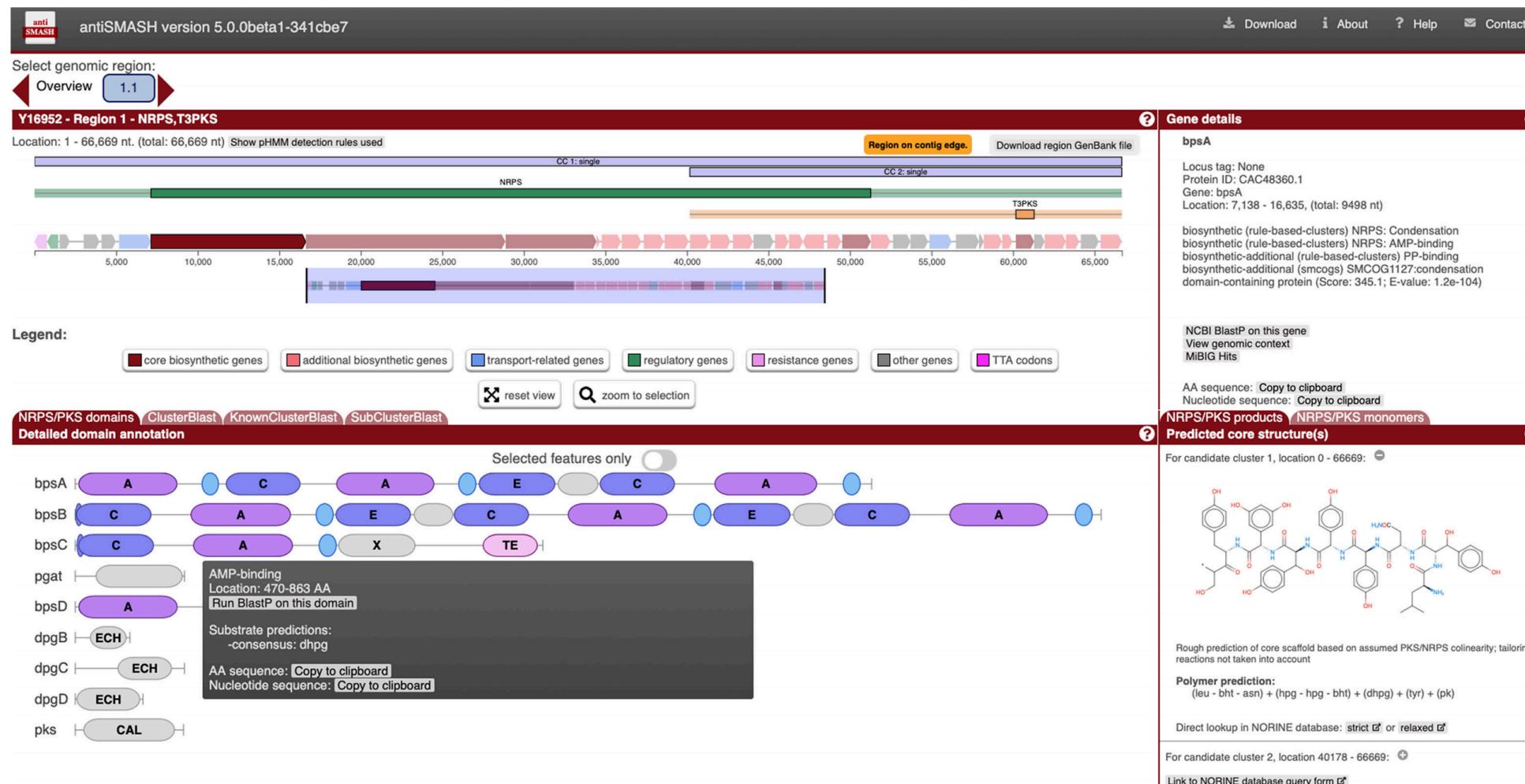
antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification

Kai Blin¹, Thomas Wolf², Marc G. Chevrette³, Xiaowen Lu⁴, Christopher J. Schwanen⁵, Satria A. Kautsar⁴, Hernando G. Suarez Duran⁴, Emmanuel L. C. de los Santos⁶, Hyun Uk Kim^{1,7}, Mariana Nave⁸, Jeroen S. Dickschat⁹, Douglas A. Mitchell^{5,10}, Ekaterina Shelest², Rainer Breitling¹¹, Eriko Takano¹¹, Sang Yup Lee^{1,7}, Tilmann Weber^{1,*} and Marnix H. Medema^{4,*}

Links to other Web-resources
antiSMASH-DB
MIBiG repository
NCBI BLAST+
NaPDoS
Norine
Output file formats
Genbank
EMBL
BiosynML
Tab-delimited text files
Input file formats
FASTA (nucleotide or protein)
FASTA + GFF3
Genbank / Genpept
EMBL
Direct download via NCBI accession number

antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline

Kai Blin¹, Simon Shaw¹, Katharina Steinke², Rasmus Villebro¹, Nadine Ziemert², Sang Yup Lee^{1,3}, Marnix H. Medema^{1,4,*} and Tilmann Weber^{1,*}



Web interface of AntiSMASH: input

antiSMASH bacterial version 

Submit Bacterial Sequence  Submit Fungal Sequence  Submit Plant Sequence  Download 

Server status: **working**

Running jobs: **10**

Queued jobs: **0**

Jobs processed: **738610**

Nucleotide input **Results for existing job**

Search a genome sequence for secondary metabolite biosynthetic gene clusters

Load sample input Open example output

Notification settings

your@email.com Email address (optional)

Data input

Upload file Get from NCBI NCBI acc # NCBI accession number of desired sequence

Detection strictness: relaxed

strict  relaxed loose

- Detects well-defined clusters containing all required parts.
- Detects partial clusters missing one or more functional parts.

Extra features All off All on

KnownClusterBlast ClusterBlast SubClusterBlast

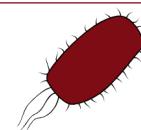
ActiveSiteFinder Cluster Pfam analysis Pfam-based GO term annotation

Submit

Please be considerate in your use of antiSMASH. Help us keep antiSMASH available for everybody by limiting yourself to 5 concurrent jobs. Need to run more? See the [antiSMASH install guide](#) for instructions for getting your own antiSMASH installation.



If you have found antiSMASH useful, please cite us.



Web interface of AntiSMASH: input

antiSMASH bacterial version 

Submit Bacterial Sequence  Submit Fungal Sequence  Submit Plant Sequence  Download

Server status: **working**

Running jobs: **8**

Queued jobs: **0**

Jobs processed: **739711**

Nucleotide input **Results for existing job**

Search a genome sequence for secondary metabolite biosynthetic gene clusters

[Load sample input](#) [Open example output](#)

Notification settings

Email address (optional)

Data input

[Upload file](#) [Get from NCBI](#) [Browse](#) **Sequence file (GenBank / EMBL / FASTA format)**

Detection strictness: relaxed

strict  loose

- Detects well-defined clusters containing all required parts.
- Detects partial clusters missing one or more functional parts.

Extra features [All off](#) [All on](#)

KnownClusterBlast ClusterBlast SubClusterBlast

ActiveSiteFinder Cluster Pfam analysis Pfam-based GO term annotation

Submit

Please be considerate in your use of antiSMASH. Help us keep antiSMASH available for everybody by limiting yourself to 5 concurrent jobs. Need to run more? See the [antiSMASH install guide](#) for instructions for getting your own antiSMASH installation.

Web interface of AntiSMASH: output

antiSMASH version 5.1.2

Select genomic region:

Overview 3.1 103.1 144.1 688.1 922.1 1097.1 1726.1 1749.1 3217.1 3502.1

Download all results
Download GenBank summary file
Download log file

Compact view

Identified secondary metabolite regions using strictness 'relaxed'

3

Region Type From To Most similar known cluster Similarity

Region 3.1 bacteriocin 4,541 16,448

103

Region Type From To Most similar known cluster Similarity

Region 103.1 betalactone 7,310 21,858

144

Region Type From To Most similar known cluster Similarity

Region 144.1 bacteriocin 1 9,782

688

Region Type From To Most similar known cluster Similarity

Region 688.1 acyl_amino_acids 1 8,815

922

Region Type From To Most similar known cluster Similarity

Region 922.1 T3PKS 1 7,415

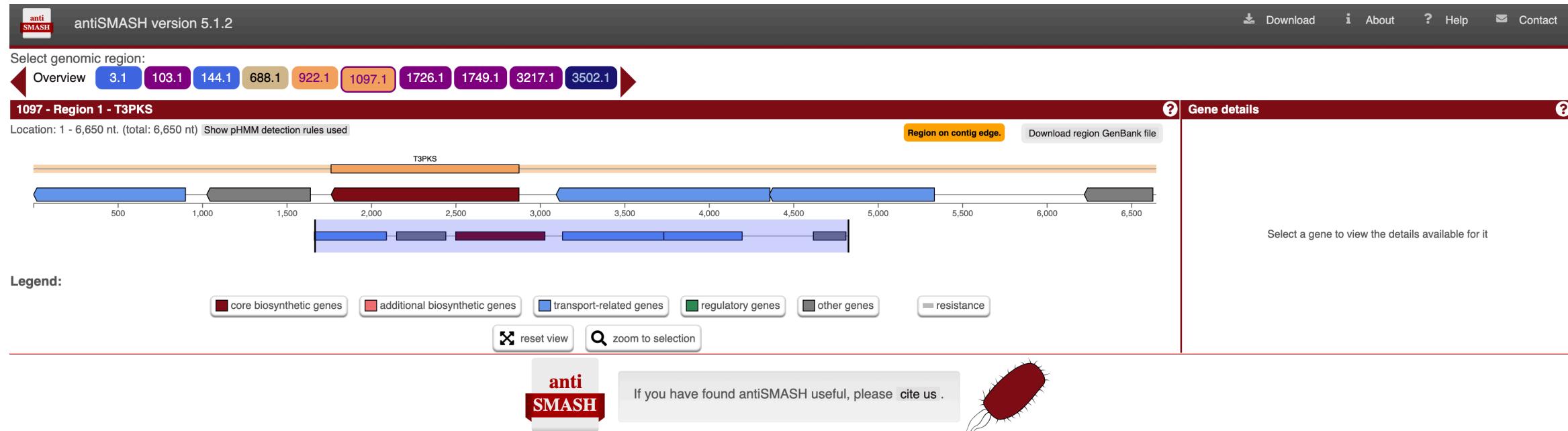
1097

Region Type From To Most similar known cluster Similarity

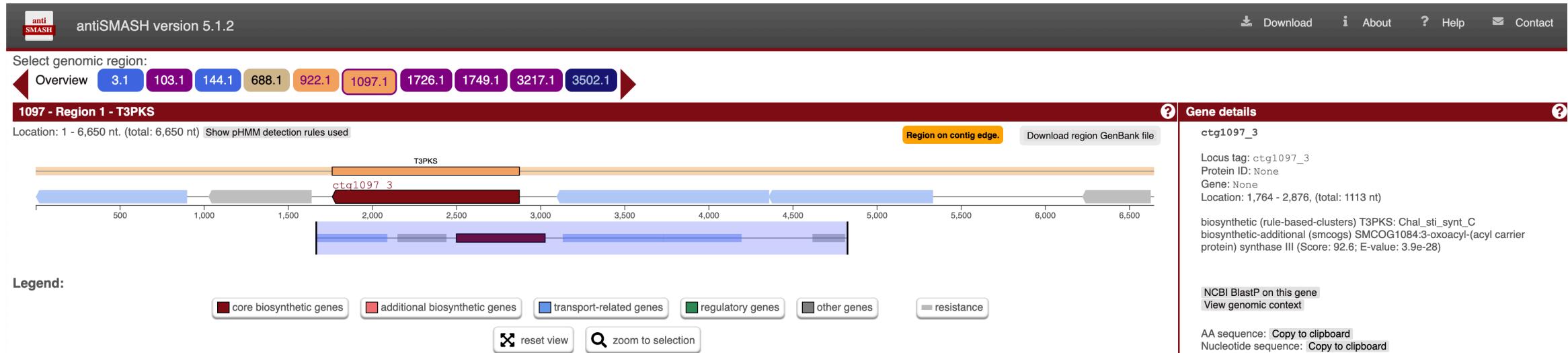
Region 1097.1 T3PKS 1 6,650

1726

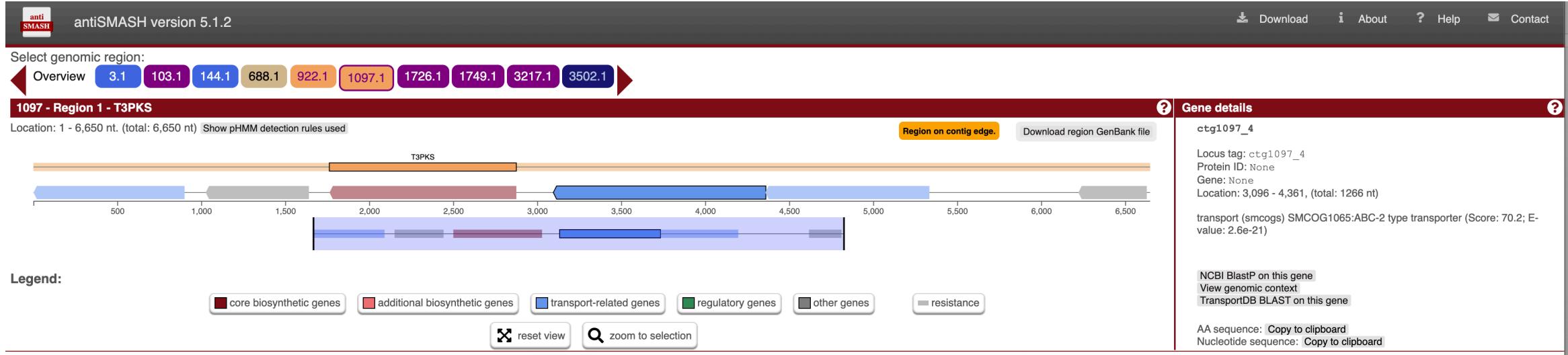
Web interface of AntiSMASH: output



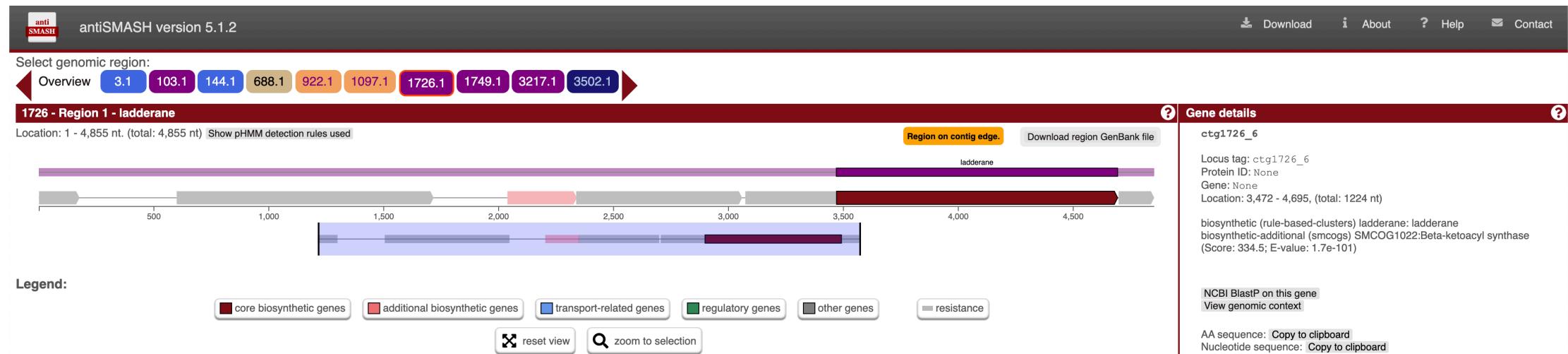
Web interface of AntiSMASH: output



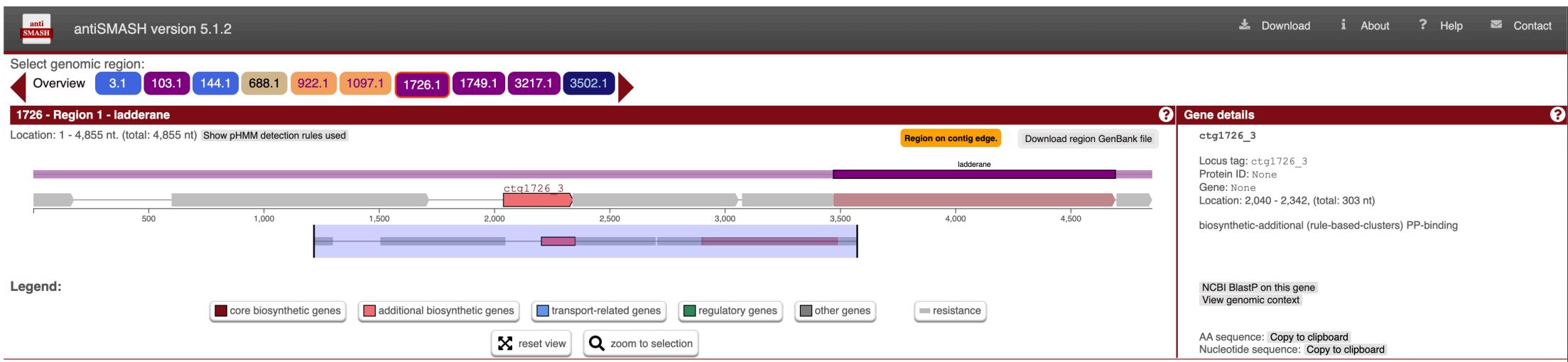
Web interface of AntiSMASH: output



Web interface of AntiSMASH: output



Web interface of AntiSMASH: output



Web interface of AntiSMASH: output

antiSMASH version 5.1.2

Select genomic region:

Overview 3.1 103.1 144.1 688.1 922.1 1097.1 1726.1 1749.1 3217.1 3502.1

Identified secondary metabolite regions using strictness 'relaxed'

3

Region Type From To Most similar known cluster Similarity

Region 3.1 bacteriocin 4,541 16,448

103

Region Type From To Most similar known cluster Similarity

Region 103.1 betalactone 7,310 21,858

144

Region Type From To Most similar known cluster Similarity

Region 144.1 bacteriocin 1 9,782

688

Region Type From To Most similar known cluster Similarity

Region 688.1 acyl_amino_acids 1 8,815

922

Region Type From To Most similar known cluster Similarity

Region 922.1 T3PKS 1 7,415

1097

Region Type From To Most similar known cluster Similarity

Region 1097.1 T3PKS 1 6,650

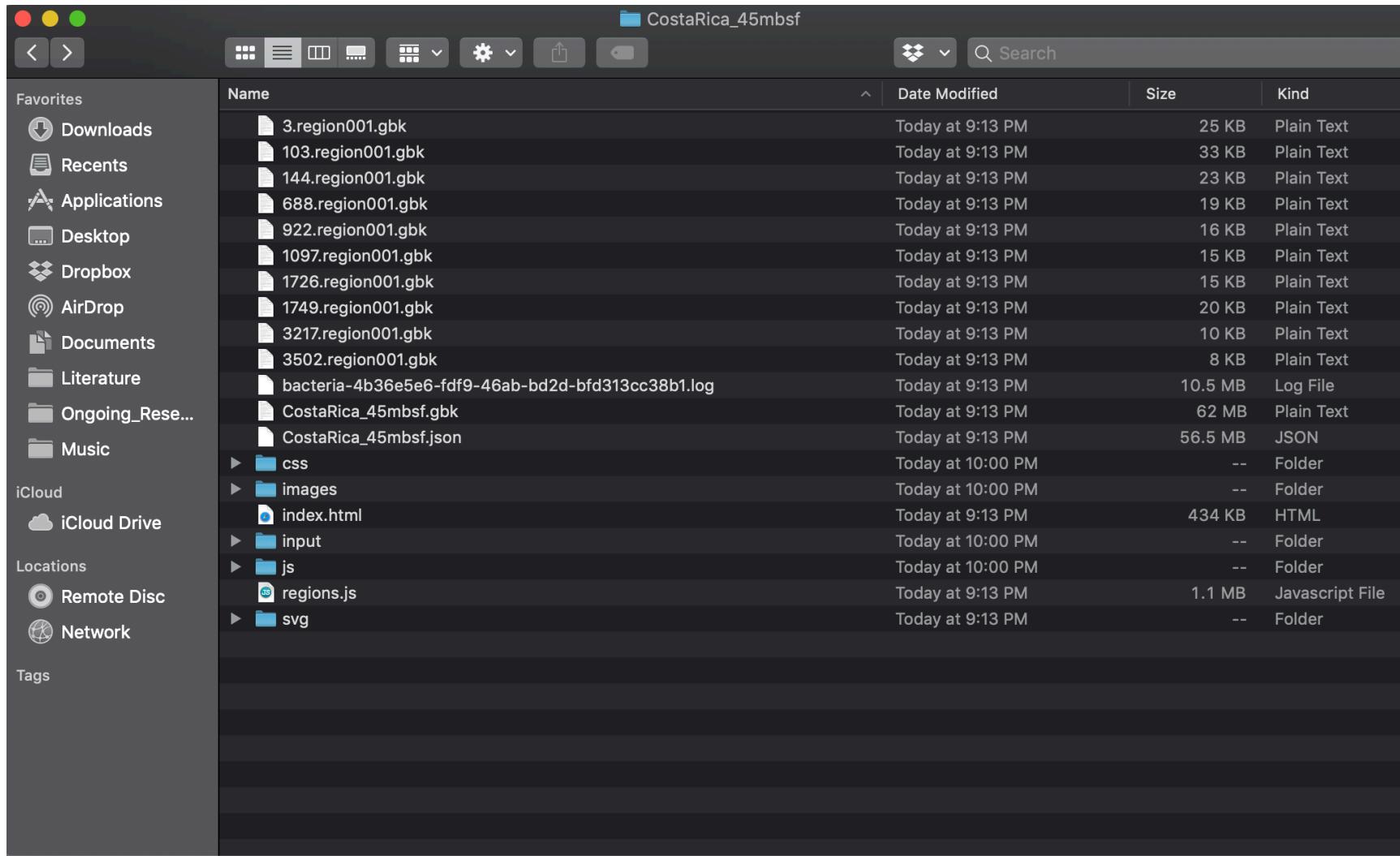
1726

Download all results
Download GenBank summary file
Download log file

About Help Contact

Compact view

Bioconda-installable standalone version



Bioconda-installable standalone version

Using Bioconda

Bioconda is a channel for the [conda](#) package manager with a focus on bioinformatics software. Once you have [bioconda installed](#), installing antiSMASH is as easy as running

```
conda create -n antismash antismash
conda activate antismash
download-antismash-databases
conda deactivate
```

Later, if you want to run antiSMASH, simply call

```
conda activate antismash
antismash my_input.gbk
```

Using Docker

If you're not on Debian, but don't want to bother with the full-blown manual installation, there are two docker images you can use.

antiSMASH standalone

This image includes all the required databases, no further configuration needed. It is a ~9 GB download at the moment. If you have docker installed, just grab the wrapper script and get going:

```
mkdir ~/bin    # not required if you already have that
curl -q https://dl.secondarymetabolites.org/releases/latest/docker-run_antismash-full > ~/bin/run_antismash
chmod a+x ~/bin/run_antismash
run_antismash <input file> <output directory> [antismash options]
```

Bioconda-installable standalone version

```
##### antiSMASH 5.1.2 #####
usage: antismash [-h] [options ..] sequence

arguments:
  SEQUENCE  GenBank/EMBL/FASTA file(s) containing DNA.

-----
Options
-----
-h, --help           Show this help text.
--help-showall      Show full lists of arguments on this help text.
-c CPUS, --cpus CPUS  How many CPUs to use in parallel. (default: 4)

Basic analysis options:
  --taxon {bacteria,fungi}
                                Taxonomic classification of input sequence. (default:
                                bacteria)

Additional analysis:
  --fullhmmer          Run a whole-genome HMMer analysis.
  --cassis              Motif based prediction of SM gene cluster regions.
  --cf-borders-only    Only annotate borders of existing clusters.
  --cf-create-clusters Find extra clusters.
  --clusterhmmer       Run a cluster-limited HMMer analysis.
  --smcog-trees        Generate phylogenetic trees of sec. met. cluster
                        orthologous groups.
  --tta-threshold TTA_THRESHOLD
                        Lowest GC content to annotate TTA codons at (default:
                        0.65).
  --cb-general         Compare identified clusters against a database of
                        antiSMASH-predicted clusters.
  --cb-subclusters    Compare identified clusters against known subclusters
                        responsible for synthesising precursors.
  --cb-knownclusters  Compare identified clusters against known gene
                        clusters from the MiBiG database.
  --ASF                Run active site finder analysis.
  --pfam2go            Run Pfam to Gene Ontology mapping module.

Output options:
  --output-dir OUTPUT_DIR
                        Directory to write results to.
  --html-title HTML_TITLE
                        Custom title for the HTML output page (default is
                        input filename).
  --html-description HTML_DESCRIPTION
                        Custom description to add to the output.

Gene finding options (ignored when ORFs are annotated):
  --genefinding-tool {glimmerhmm,prodigal,prodigal-m,none,error}
                        Specify algorithm used for gene finding: GlimmerHMM,
                        Prodigal, Prodigal Metagenomic/Anonymous mode, or
                        none. The 'error' option will raise an error if
                        genefinding is attempted. The 'none' option will not
                        run genefinding. (default: error).
  --genefinding-gff3 GFF3_FILE
                        Specify GFF3 file to extract features from.

(antismash) [agarber4@agave3:-]$ antismash --help
```



Onto the Jupyter Binder Tutorial: Standalone AntiSMASH

