



UNIVERSITY OF HELSINKI
FACULTY OF AGRICULTURE AND FORESTRY

Detection and analysis of secondary metabolites gene clusters

Tânia Keiko Shishido Joutsen
tania.shishido@helsinki.fi

MBDP105
31.03.2022



Cyanobacterial natural products

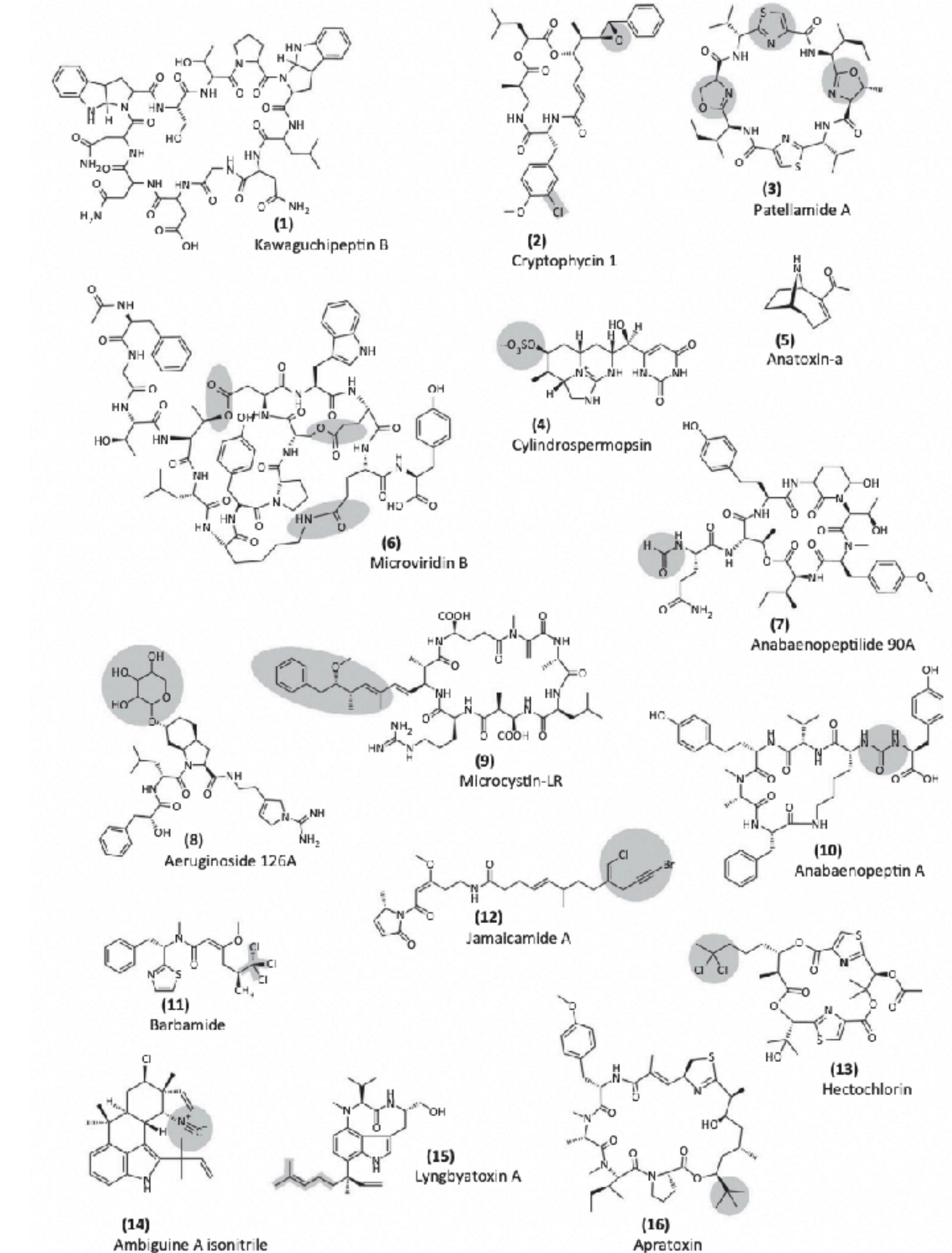
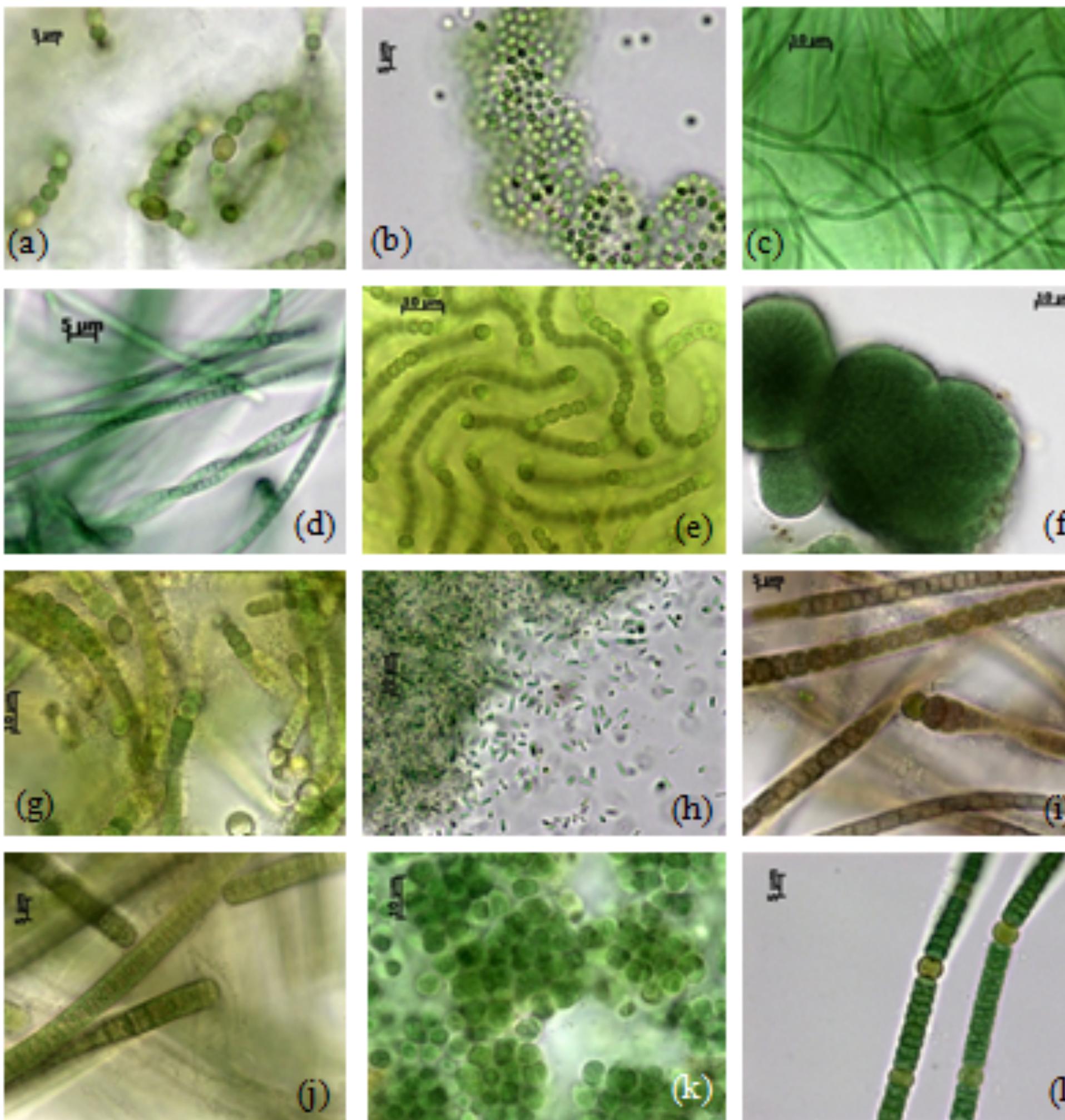
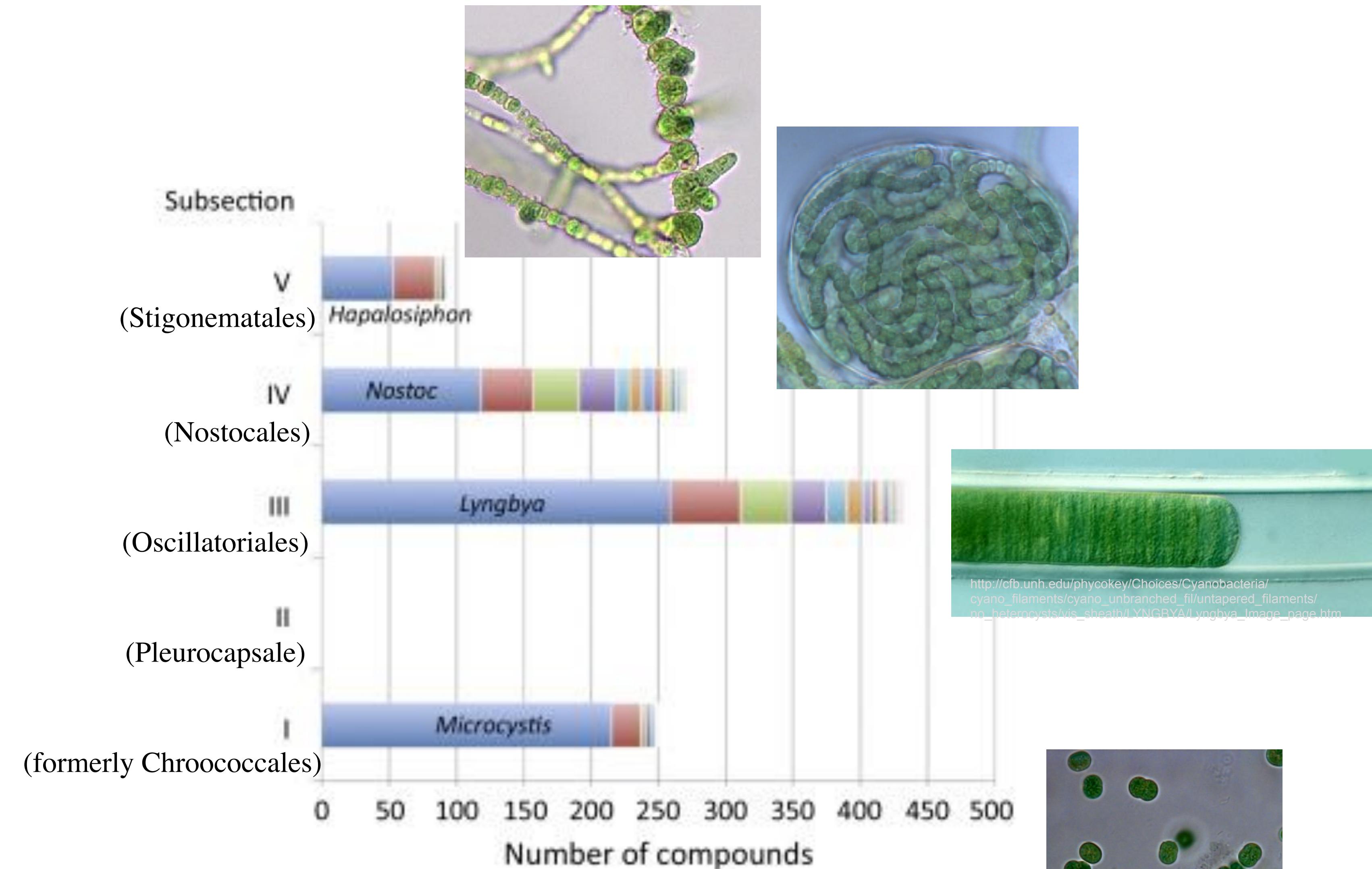
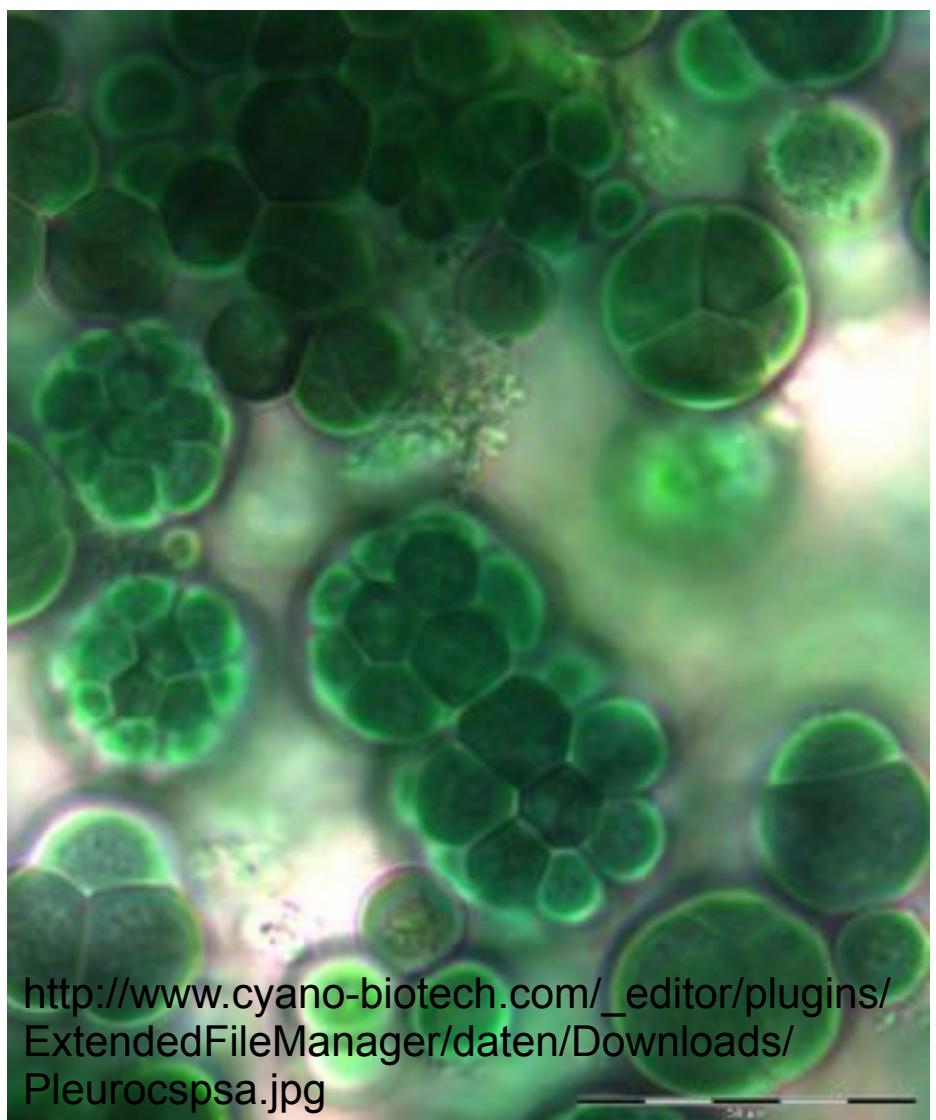


Figure 1. A Selection of Cyanobacterial Natural Products. Characteristic structural features are highlighted in gray.

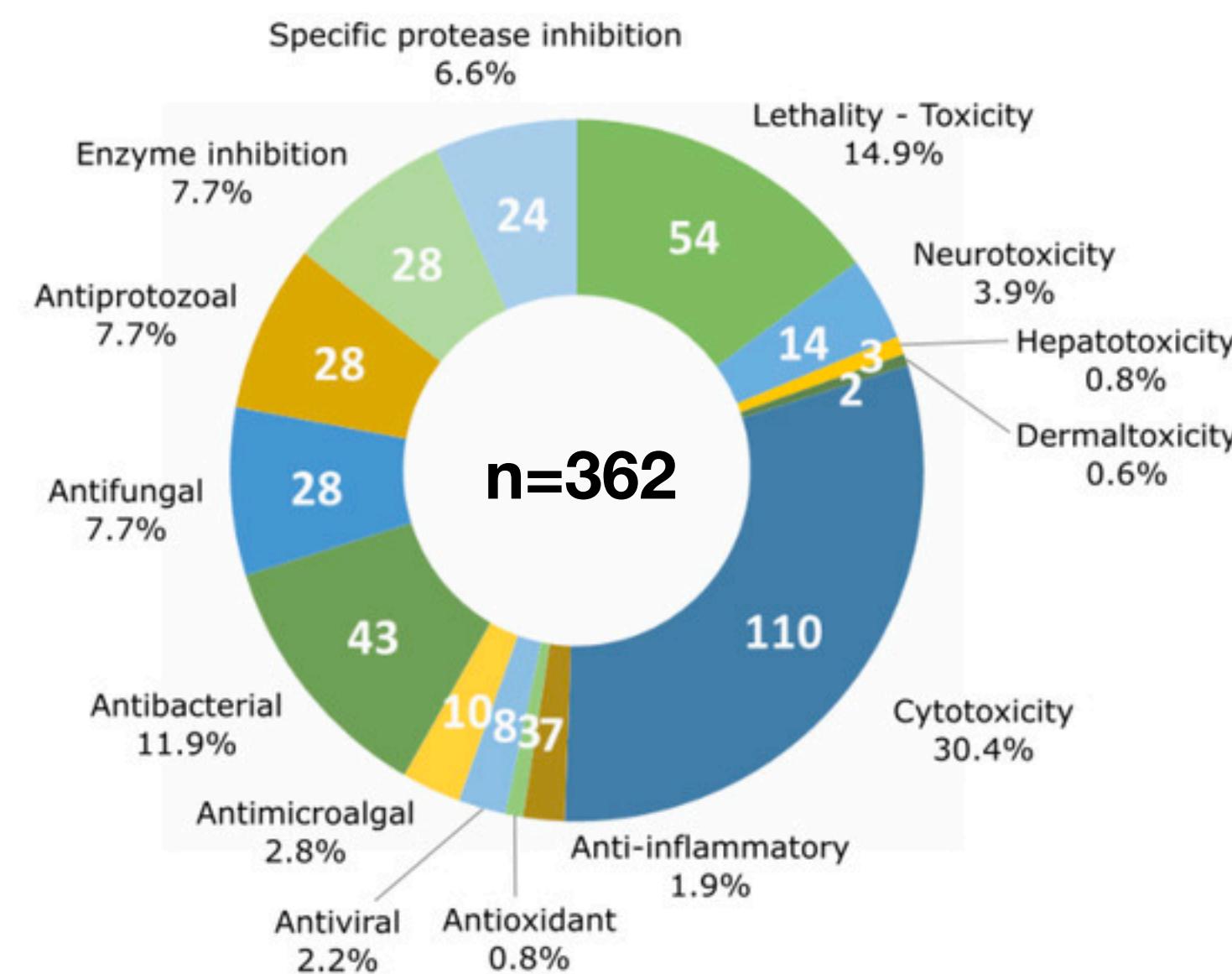
Cyanobacterial natural products



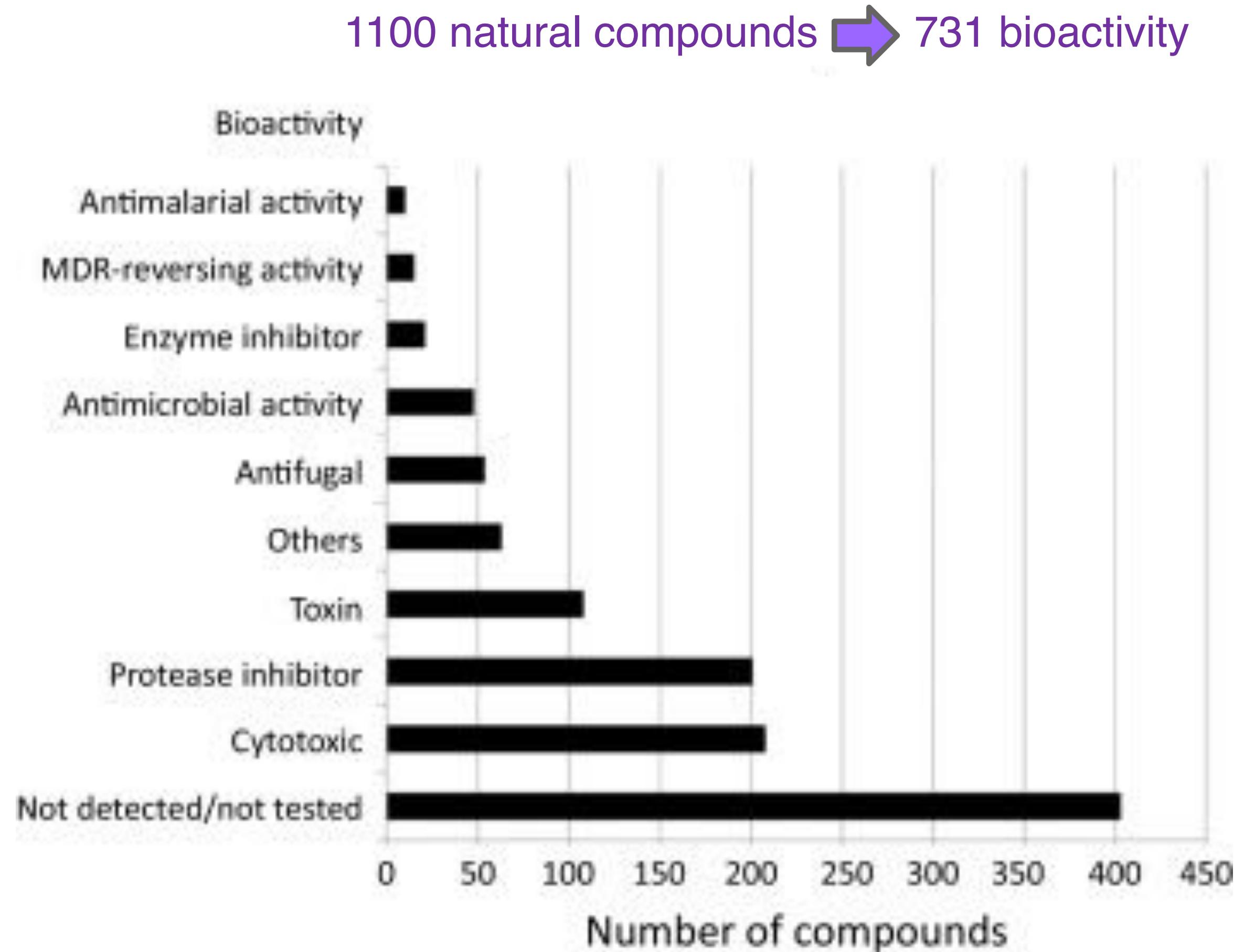
Dittmann et al. 2015 Trends Microbiol.



Cyanobacterial natural products



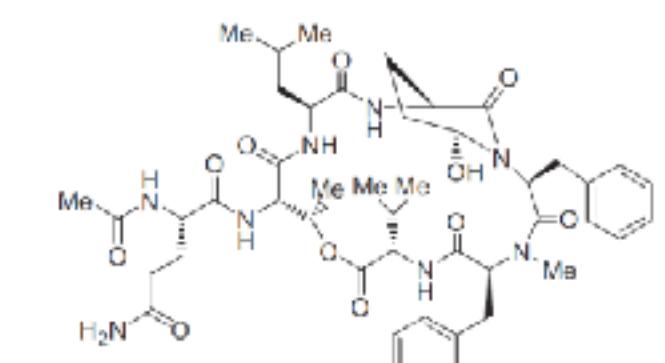
Demay et al. 2019 Mar Drugs



Trends in Microbiology

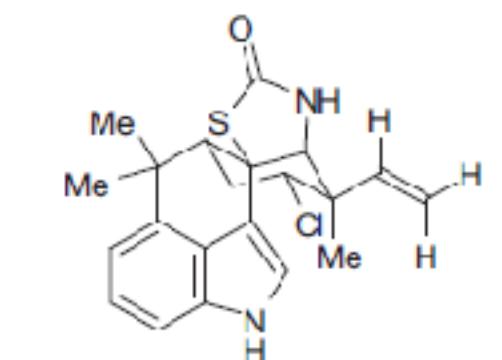
Dittmann et al. 2015 Trends Microbiol.

Protease inhibitor



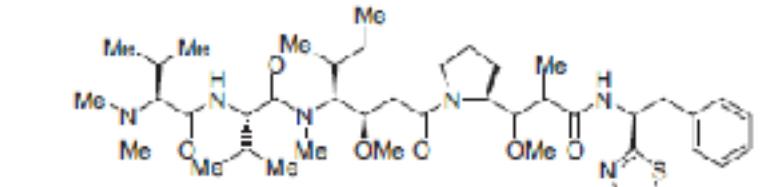
Cyanopeptolin 954

Antibacterial



Hapalindole T

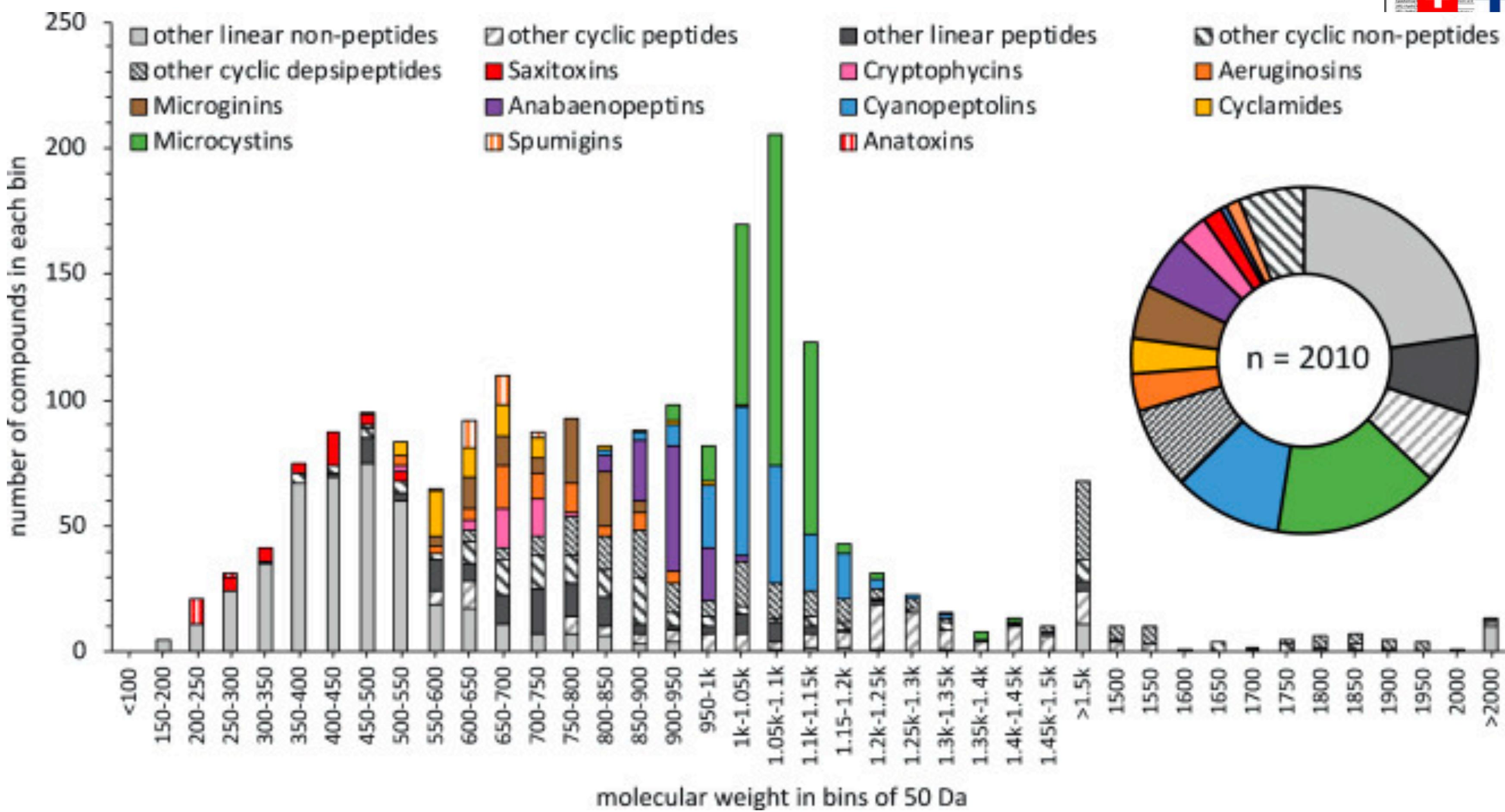
Anticancer



Dolastatin 10



Cyanobacterial natural products



CyanoMetDB secondary metabolites

Open Access Database

private lists > 850 references >2000 structural codes to be included in norman MetFrag PubChem npatlas CompTox

CC1/C=C(C)/C=C/C(=O)C(C(=O)N(C)C)=OCCC(N(C)C)OC(=O)C(C(=O)N(C)C)C1=O
NC(=O)N1=OCC(C)CC2=CC=C(C=C2)C1=O

Genetic diversity of natural product biosynthetic pathways:

- recombination
- gene deletion
- the gain and loss of tailoring enzymes
- horizontal gene transfer



← → C ncbi.nlm.nih.gov/genome/browse#!/overview/cyanobacteria

Log in

NIH National Library of Medicine
National Center for Biotechnology Information

i New Genome Table
Try our new [Genome page](#) and use the feedback button to let us know what you think

Genome > Genome Information by Organism

cyanobacteria

X Q Search

Download Reports from FTP site

Overview (668); Prokaryotes (3552); Plasmids (511)

▼ Filters [Download](#)

Choose Columns

Page 1 of 14 | 50

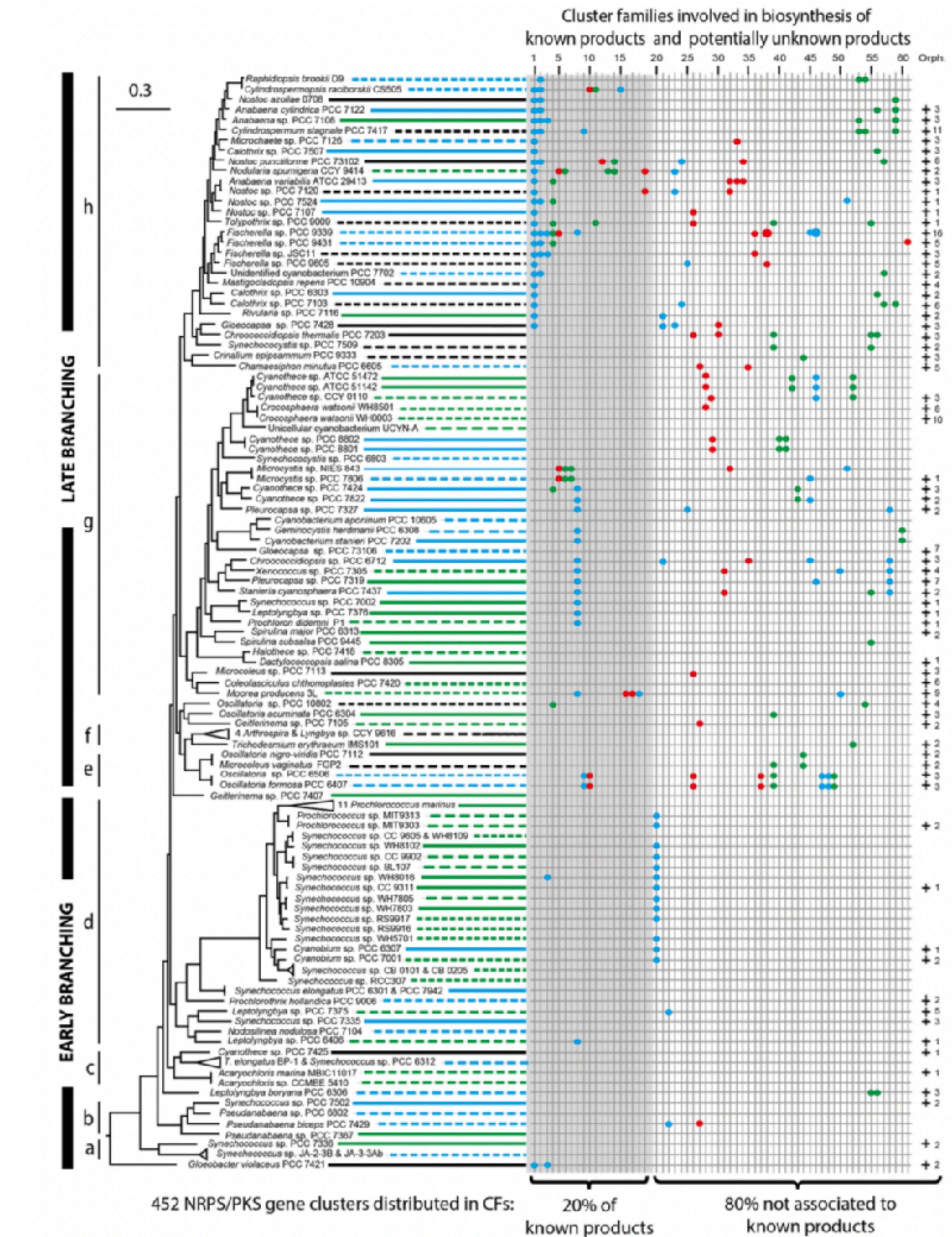
View 1 - 50 of 668

#	Organism Name	Organism Groups	Size(Mb)	Chromosomes	Organelles	Plasmids	Assemblies
1	Acaryochloridaceae cyanobacterium CSU_3_4	Bacteria;Terrabacteria group;Cyanobacteria/Melainabacteria group	2.74569	-	-	-	1
2	Acaryochloridaceae cyanobacterium CSU_5_19	Bacteria;Terrabacteria group;Cyanobacteria/Melainabacteria group	1.28234	-	-	-	1
3	Acaryochloridaceae cyanobacterium RL_2_7	Bacteria;Terrabacteria group;Cyanobacteria/Melainabacteria group	1.46145	-	-	-	1
4	Acaryochloridaceae cyanobacterium RU_4_10	Bacteria;Terrabacteria group;Cyanobacteria/Melainabacteria group	4.48368	-	-	-	1
5	Acaryochloridaceae cyanobacterium SU_2_1	Bacteria;Terrabacteria group;Cyanobacteria/Melainabacteria group	3.0211	-	-	-	1
6	Acaryochloris	Bacteria;Terrabacteria group;Cyanobacteria/Melainabacteria group	7.87548	1	-	10	5
7	Acaryochloris marina	Bacteria;Terrabacteria group;Cyanobacteria/Melainabacteria group	8.3616	1	-	9	2
8	Acaryochloris thomasi RCC1774	Bacteria;Terrabacteria group;Cyanobacteria/Melainabacteria group	5.93184	-	-	-	1
9	Aetokthonos hydrillicola	Bacteria;Terrabacteria group;Cyanobacteria/Melainabacteria group	9.01908	-	-	-	3
10	Aliterella atlantica CENA595	Bacteria;Terrabacteria group;Cyanobacteria/Melainabacteria group	5.2657	-	-	-	1
11	Alkalinema	Bacteria;Terrabacteria group;Cyanobacteria/Melainabacteria group	6.46103	-	-	-	4
12	Allocoleopsis franciscana	Bacteria;Terrabacteria group;Cyanobacteria/Melainabacteria group	7.96651	1	-	8	1
13	Amazonocrinis nigriterrae	Bacteria;Terrabacteria group;Cyanobacteria/Melainabacteria group	8.25647	-	-	-	1
14	Anabaena	Bacteria;Terrabacteria group;Cyanobacteria/Melainabacteria group	8.56696	2	-	3	26
15	Anabaena azotica	Bacteria;Terrabacteria group;Cyanobacteria/Melainabacteria group	8.04817	-	-	-	1
16	Anabaena catenula	Bacteria;Terrabacteria group;Cyanobacteria/Melainabacteria group	6.23438	-	-	-	1
17	Anabaena cylindrica	Bacteria;Terrabacteria group;Cyanobacteria/Melainabacteria group	7.06328	1	-	6	5
18	Anabaena lutea	Bacteria;Terrabacteria group;Cyanobacteria/Melainabacteria group	6.05942	-	-	-	1
19	Anabaena minutissima	Bacteria;Terrabacteria group;Cyanobacteria/Melainabacteria group	7.0652	-	-	-	1



Cyanobacterial natural products biosynthetic pathway

89 cyanobacterial genomes - 452 NRPS and PKS gene clusters



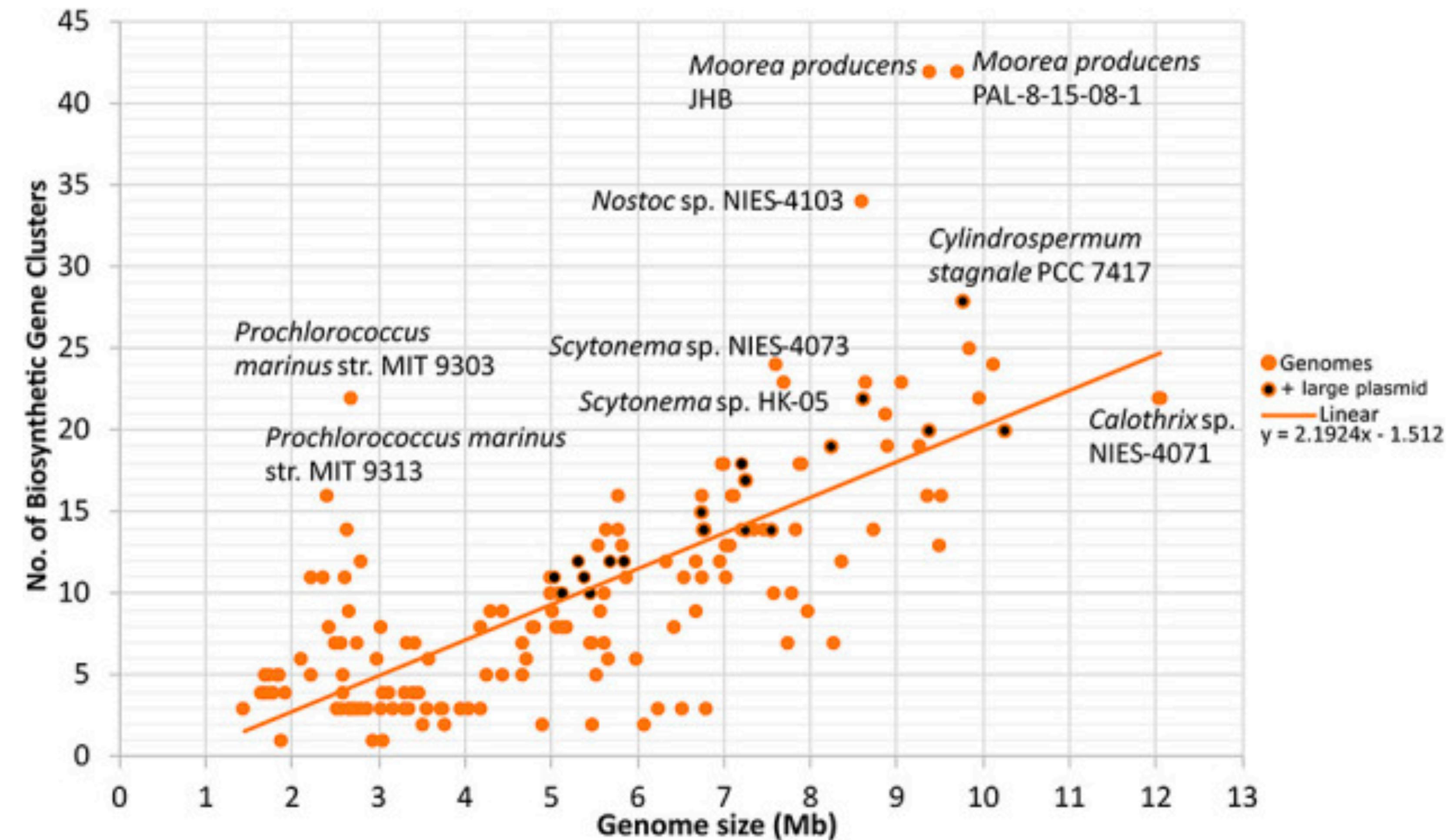


Genome mining

BGC with unknown products & products without known BGC

5-6% secondary metabolites production

Calteau et al 2014, BMC Genomics





Genome mining

Cyanobacteria are recalcitrant to genetic manipulation

Small amount produced

Heterologous expression

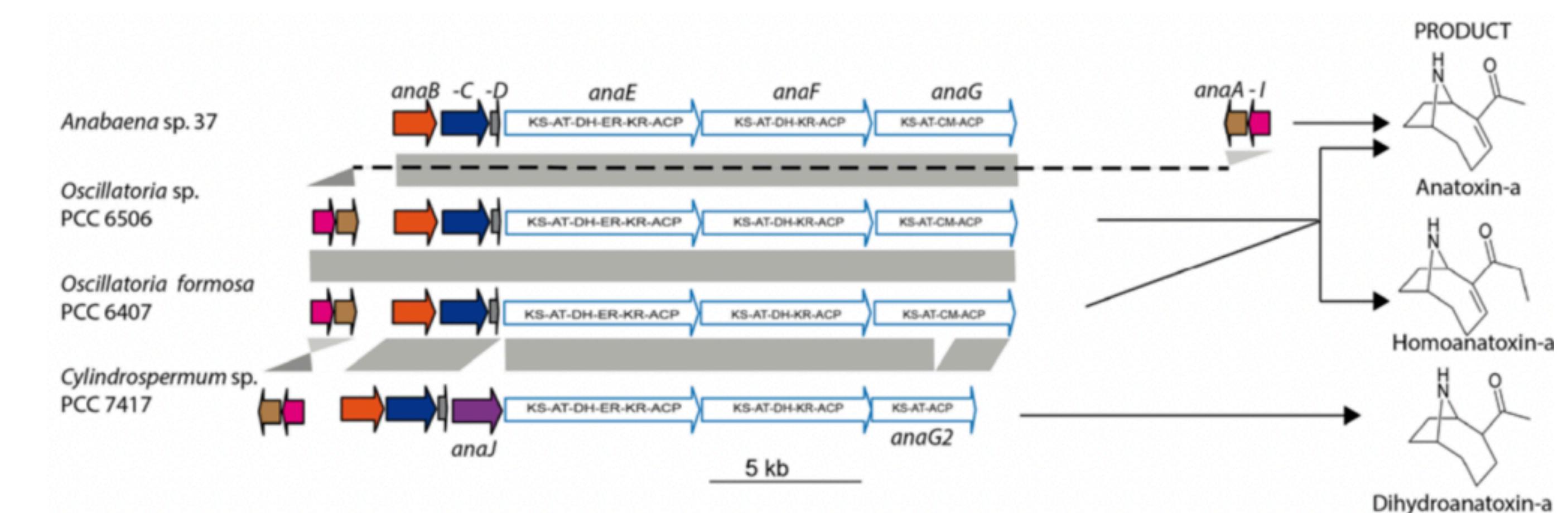


Figure 2 Anatoxin-a biosynthetic gene cluster (CF-9) and produced variants of in PCC 7417. Anatoxin-a pathway identified in the genome of *Cylindrospermum* sp. PCC 7417 compared to homologous gene clusters from anatoxin-a producing cyanobacteria [8,26,27]. Genes with corresponding functions and domain organization are colored the same and connected by grey areas: *anaA*, proline adenylation, *anaB*, proline deshydrogenase, *anaC*, Type II thioesterase; *anaD*, acyl carrier protein, *anaE*, *anaF* and *anaG* are modular type I polyketide synthase with KS, β-ketoacyl synthase; AT, acyltransferase, KR, ketoacyl reductase; ACP, acyl carrier protein; DH, dehydratase; ER, enoyl reductase and CM, C-methyltransferase. The cyclase, named here *anaI*, is systematically associated to the anatoxin-a pathway. In addition, the last PKS *anaG2* in PCC 7417 lacks the methyltransferase, and an oxidoreductase *anaJ* was detected. The transposase is present in the surrounding of the cluster only in PCC 6506. The detection of dihydroanatoxin-a from PCC 7417 is presented in Additional file 2: Figure S2.

Calteau et al 2014, BMC Genomics

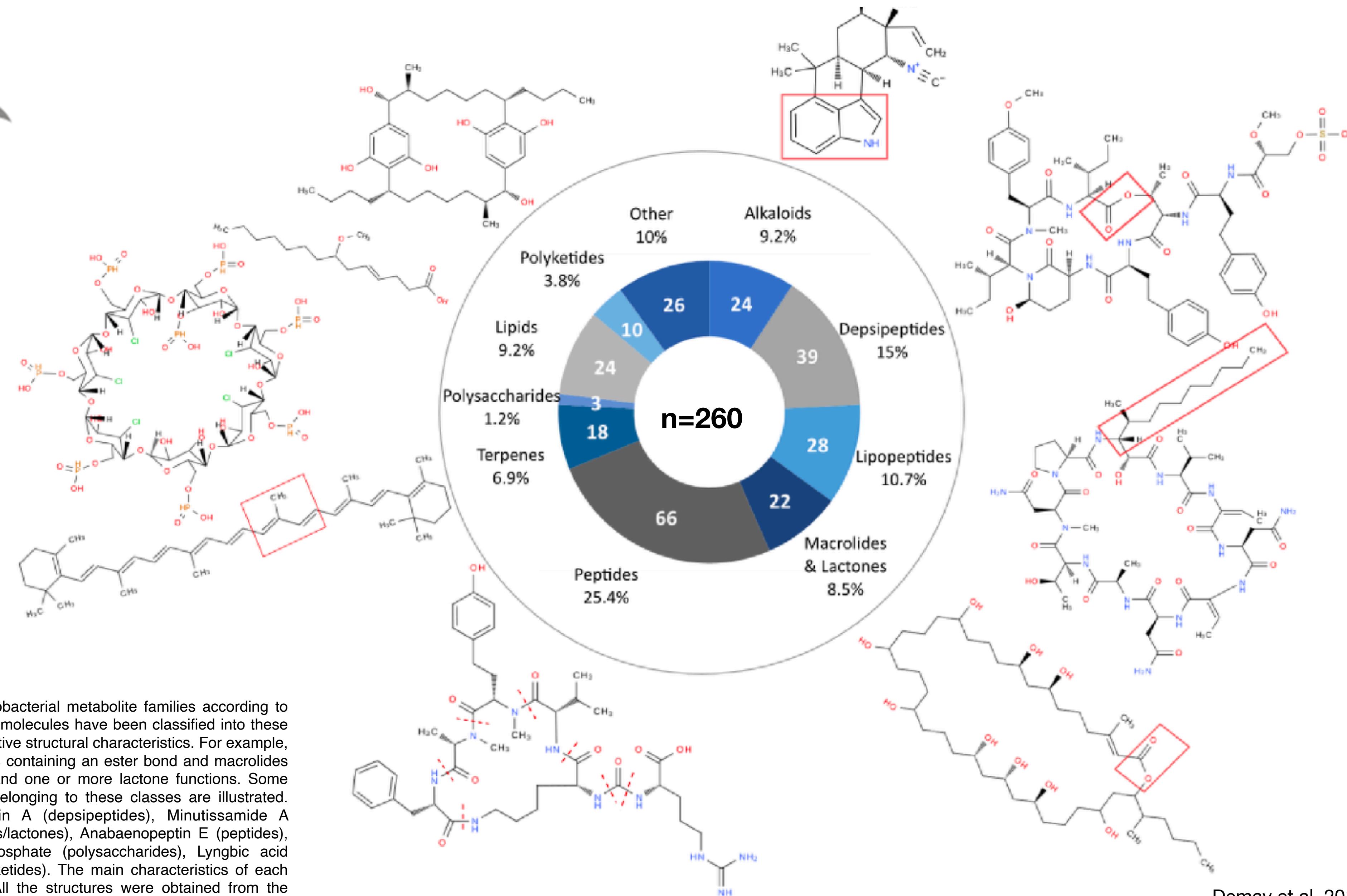
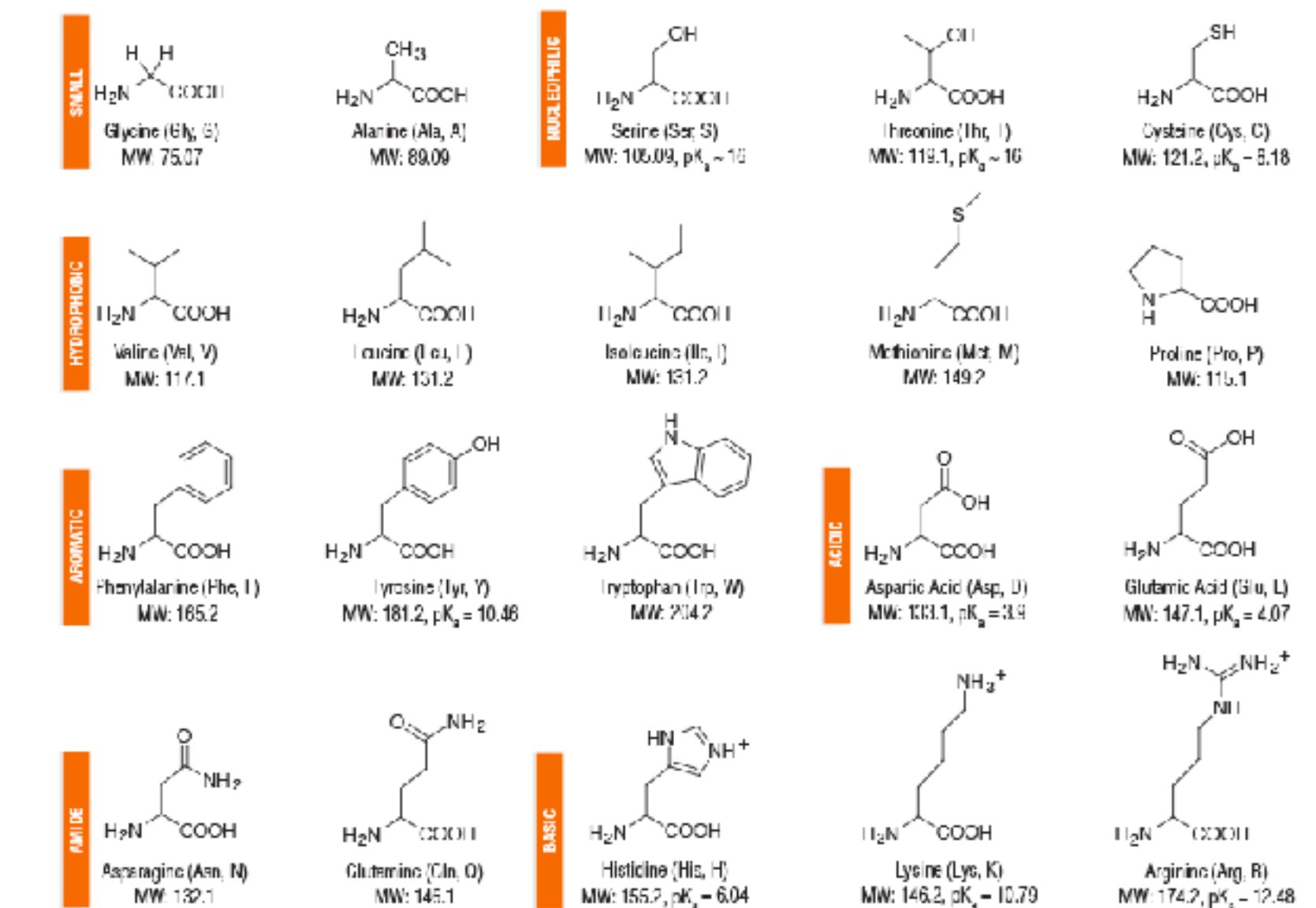
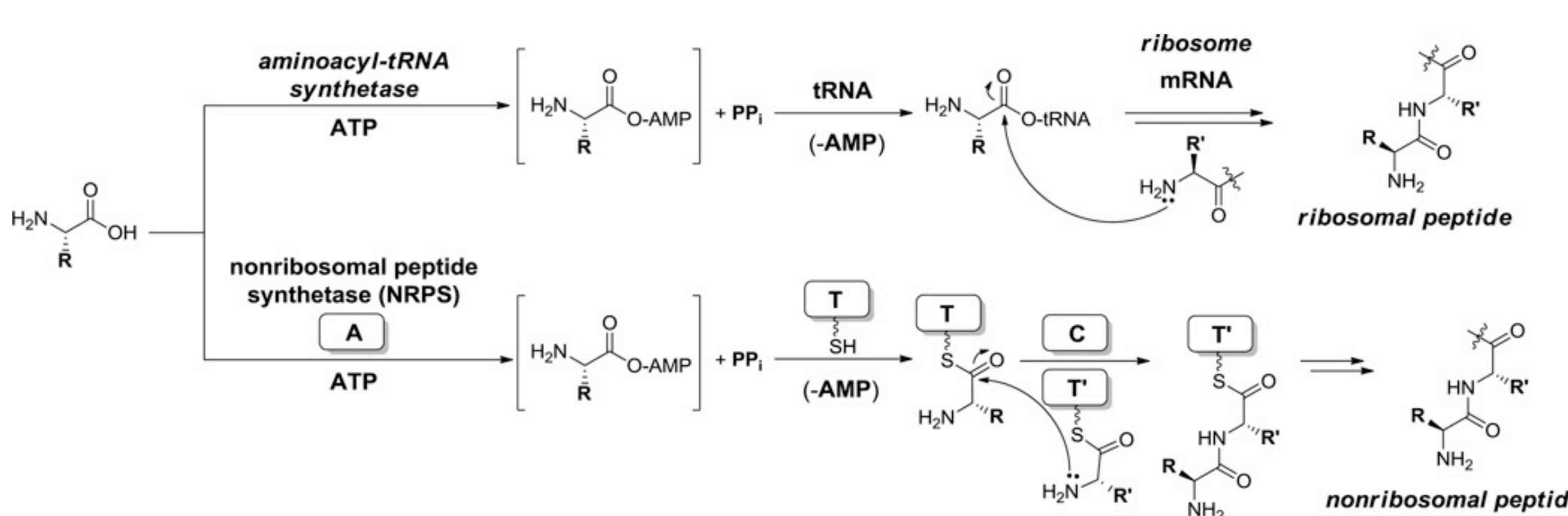


Figure 3. Classification of the 260 cyanobacterial metabolite families according to their respective chemical classes. All the molecules have been classified into these different classes according to their respective structural characteristics. For example, the depsipeptides are a class of peptides containing an ester bond and macrolides are molecules exhibiting a macrocycle and one or more lactone functions. Some examples of cyanobacterial molecules belonging to these classes are illustrated. Hapalindole A (alkaloids), Oscillapeptin A (depsipeptides), Minutissamide A (lipopeptides), Caylobolide B (macrolides/lactones), Anabaenopeptin E (peptides), β-carotene (terpenes), Cyclodextrin phosphate (polysaccharides), Lyngbic acid (lipids), and Cylindrocyclophane A (polyketides). The main characteristics of each chemical class are highlighted in red. All the structures were obtained from the ChEMBL Database

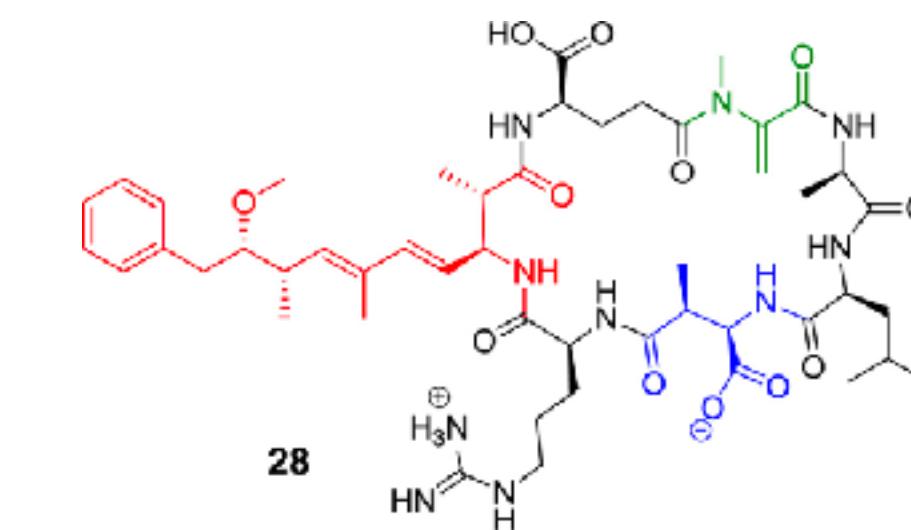


Nonribosomal Peptides and Polyketides



<https://international.neb.com/tools-and-resources/usage-guidelines/amino-acid-structures>

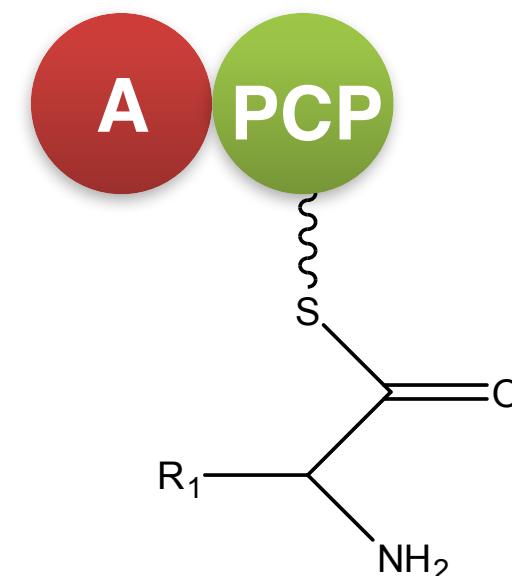
Most of the almost 500 of the nonproteinogenic amino acids are produced by a small number of enzymatic reactions, either from preexisting proteinogenic amino acids or by *de novo* construction.



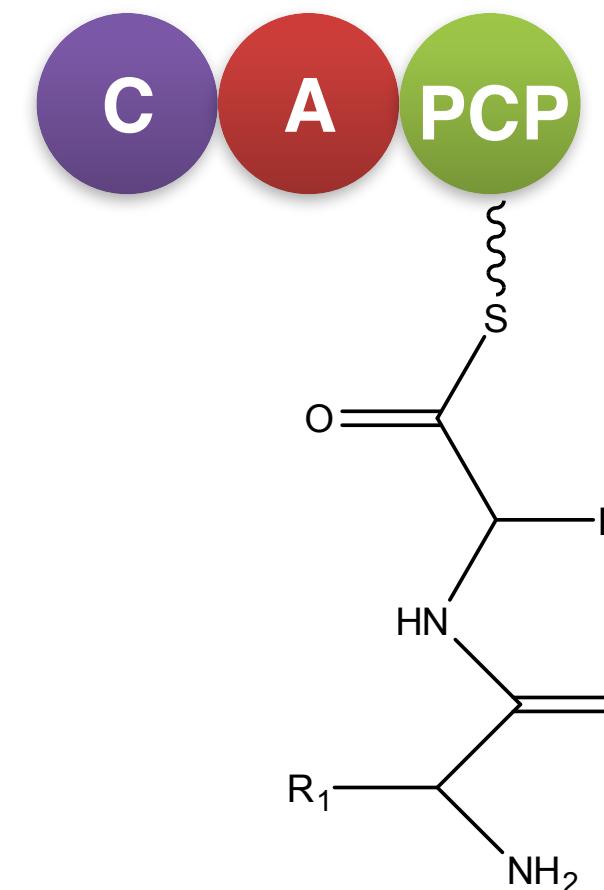


Nonribosomal biosynthetic enzymes (NRPS)

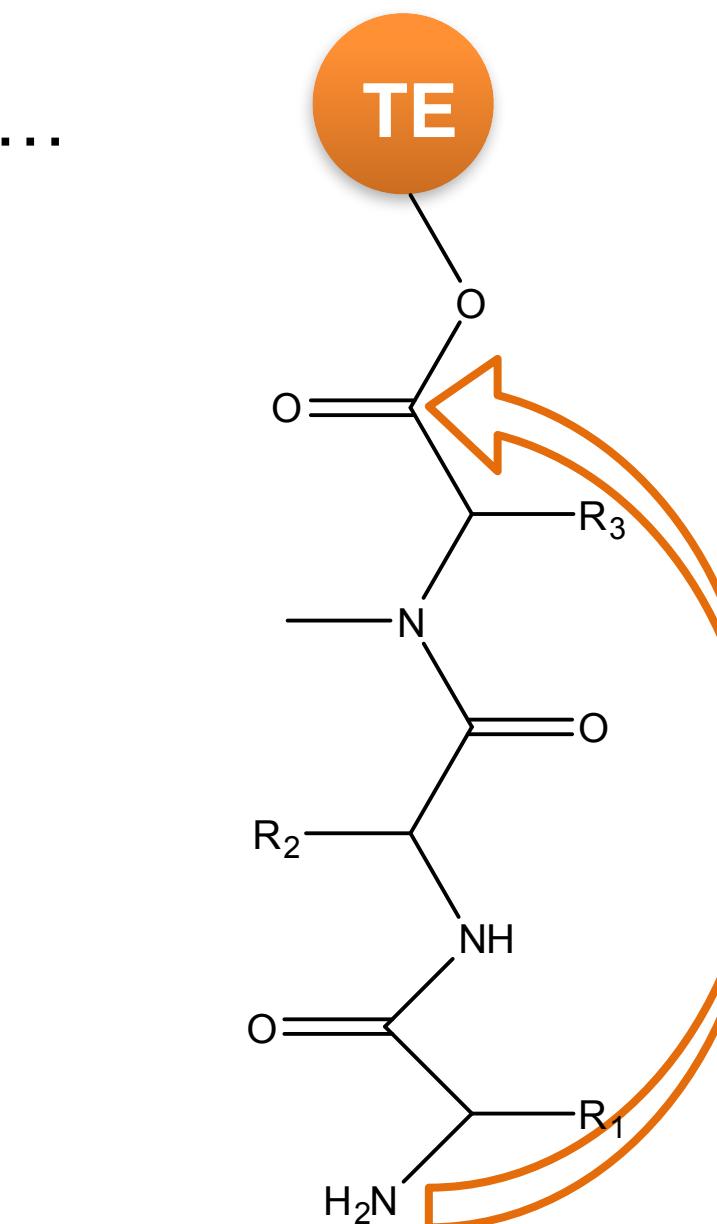
Module 1
Initiation



Module 2
Elongation

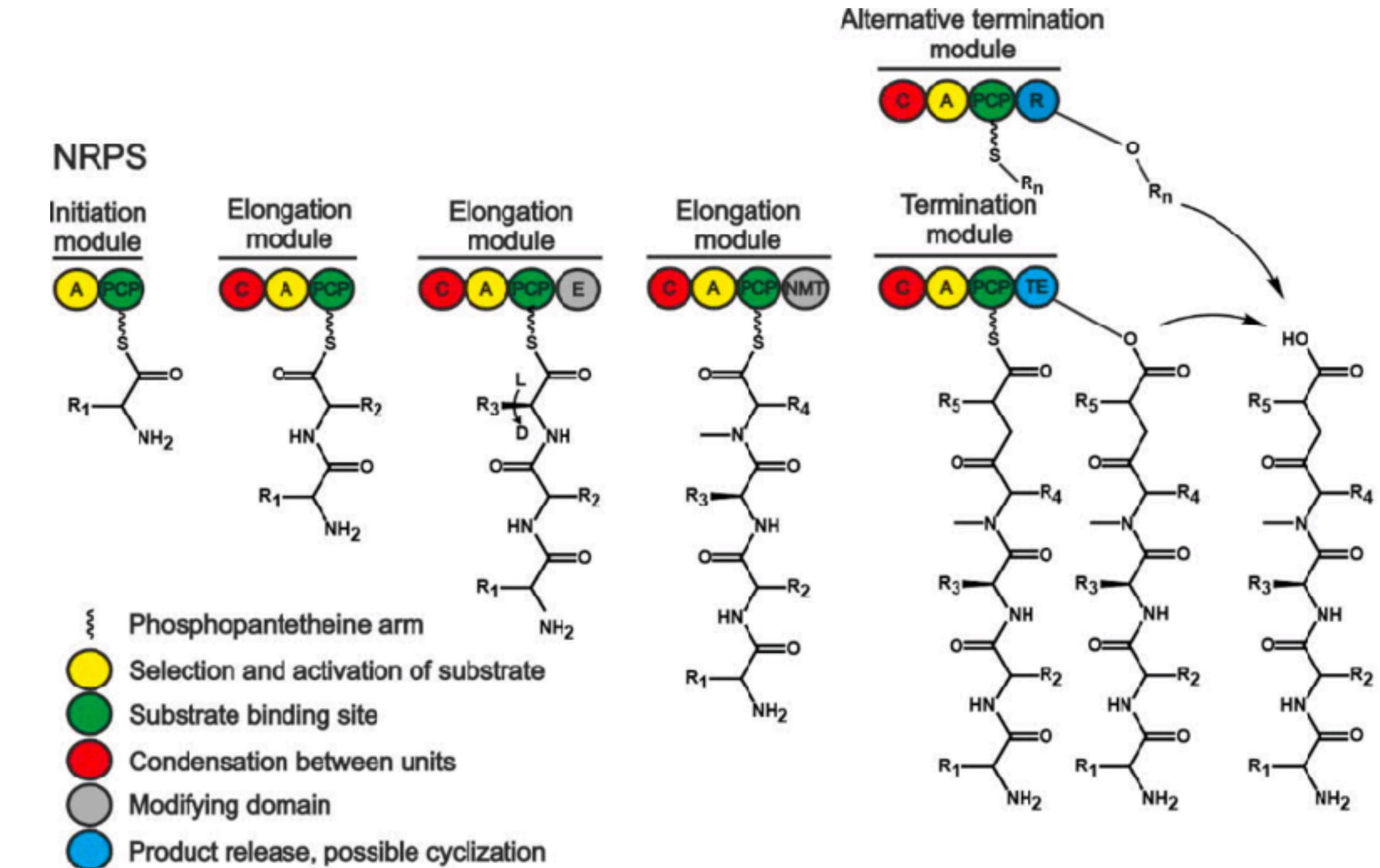


Module n

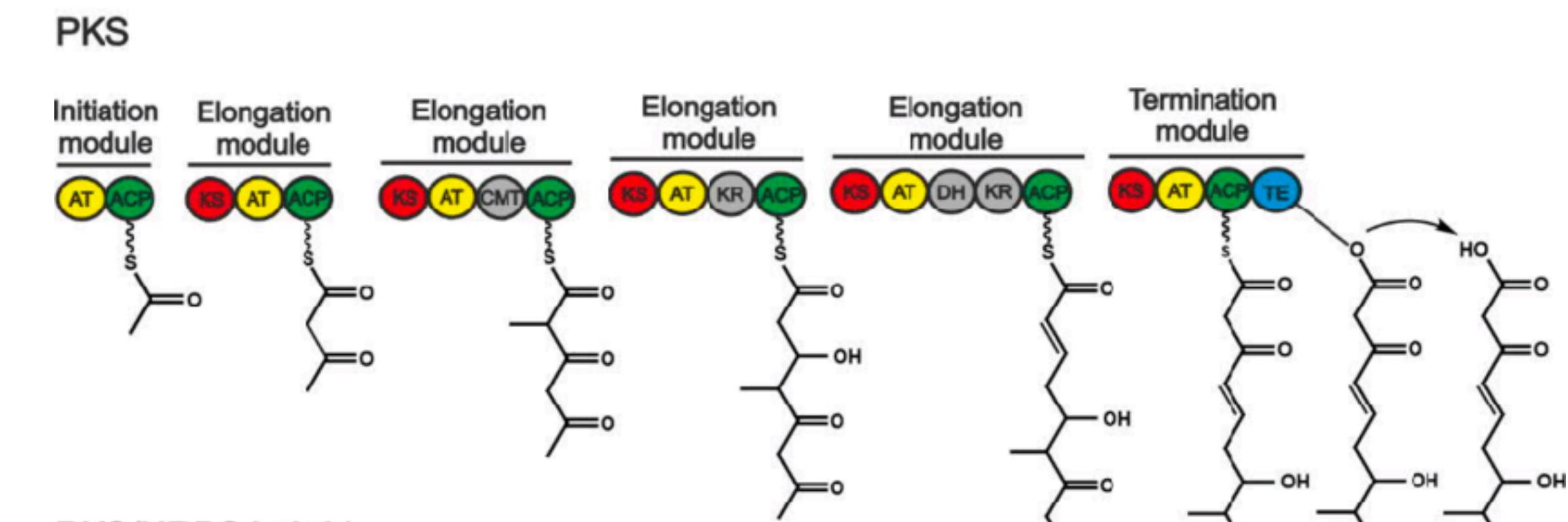




Nonribosomal peptides synthetase (NRPS)



Polyketide synthase (PKS)



Hybrid NRPS-PKS

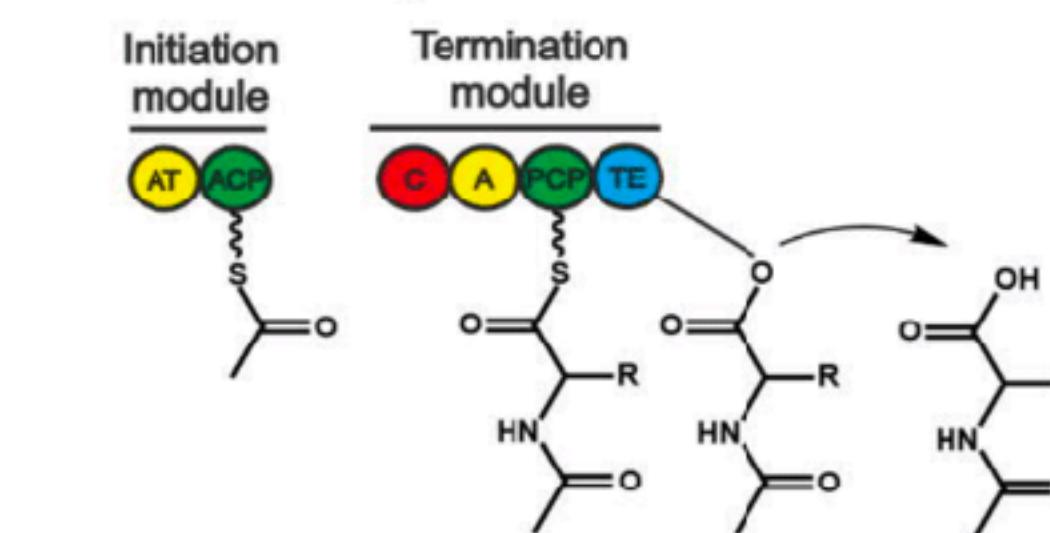


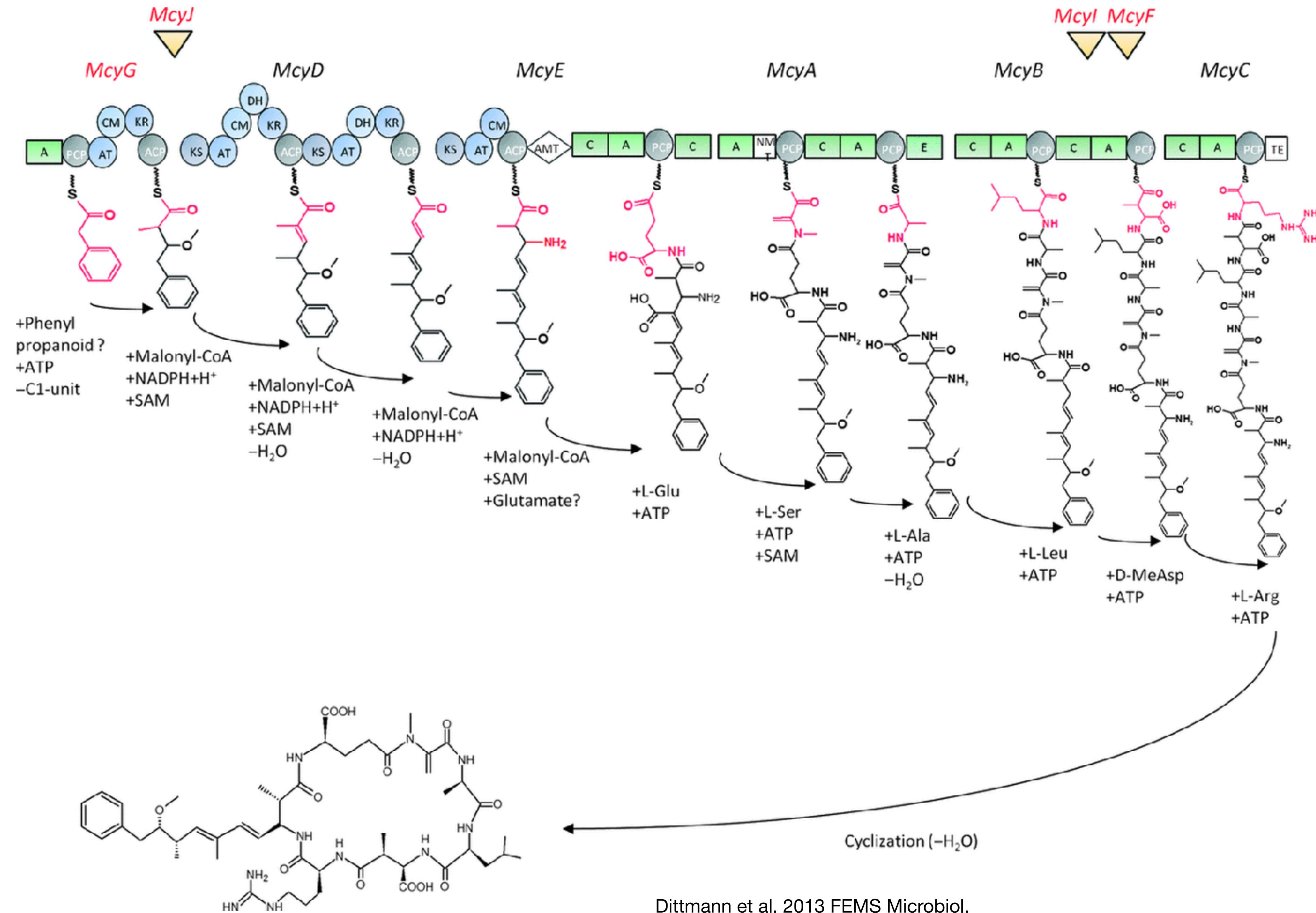
Table 1. Nonribosomal Peptides, Polyketides, Ribosomal Peptides, Indole Alkaloids, and Terpene Biosynthetic Gene Clusters from Cyanobacteria

Compound	Gene Clusters	Producer Strain	Compound Type	Size (kb)	Lineage ^a	Year	Refs
Apratoxin	?	<i>Lyngbya bouillonii</i>	NRPS-PKS	57.4	III	2011	[17]
Barbamide	AF516145	<i>Lyngbya majuscula</i> 19L	NRPS-PKS	26	III	2002	[18]
Cryptophycin	EF159954	<i>Nostoc</i> sp. ATCC 53789	NRPS-PKS	40.3	IV	2006	[21]
Curacin A	AY652953	<i>Lyngbya majuscula</i> 19L	NRPS-PKS	64	III	2004	[9]
Cylindrospermopsin	EU140798	<i>Cylindrospermopsis raciborskii</i> AWT205	NRPS-PKS	43	IV	2008	[29]
Hectochlorin	AY974560	<i>Lyngbya majuscula</i> JHB	NRPS-PKS	38	III	2007	[16]
Jamaicamide	AY522504	<i>Lyngbya majuscula</i> JHB	NRPS-PKS	58	III	2004	[22]
Microcystin	AF183408	<i>Microcystis aeruginosa</i> PCC 7806	NRPS-PKS	55	I	2000	[10]
Nodularin	AY210783	<i>Nodularia spumigena</i> NSOR10	NRPS-PKS	48	IV	2004	[25]
Nostocyclopeptide	AY167420	<i>Nostoc</i> sp. ATCC53789	NRPS-PKS	33	IV	2004	[13]
Nostophycin	JF430079	<i>Nostoc</i> sp. 152	NRPS-PKS	45.1	IV	2011	[27]
Puwainaphycins	KM078884	<i>Cylindrospermum alatosporum</i> CCALA 988	NRPS-PKS	56.7	IV	2014	[12]
Aeruginosin	AM778955	<i>Microcystis aeruginosa</i> PCC 7806	NRPS	27	I	2009	[24]
Anabaenopeptilide	AJ269505	<i>Anabaena</i> sp. 90	NRPS	29	IV	2000	[23]
Anabaenopeptin	GU174493	<i>Anabaena</i> sp. 90	NRPS	32	IV	2010	[15]
Hassallidins	KJ502174	<i>Anabaena</i> sp. SYKE 748A	NRPS	59	IV	2014	[11]
Lyngbyatoxin	AY588942	<i>Lyngbya majuscula</i>	NRPS	11.3	III	2004	[8]
Microginin	AM990464	<i>Planktothrix agardhii</i> NIVA-CYA 126/8	NRPS	20.5	III	2009	[63]
Nostopeptolide	AF204805	<i>Nostoc</i> sp. GSV224	NRPS	40	IV	2003	[28]
Spumigin	CP007203	<i>Nodularia spumigena</i> CCY 9414	NRPS	21	IV	2009	[65]
Anatoxin-a	FJ477836	<i>Oscillatoria</i> sp. PCC 6506	PKS	29	III	2009	[30]
Nosperin	JQ975876	<i>Nostoc</i> sp. 'Peltigera membranacea' cyanobiont'	PKS	22.4	IV	2013	[33]



<https://www.lgsonic.com/pt-br/cianobacterias/>

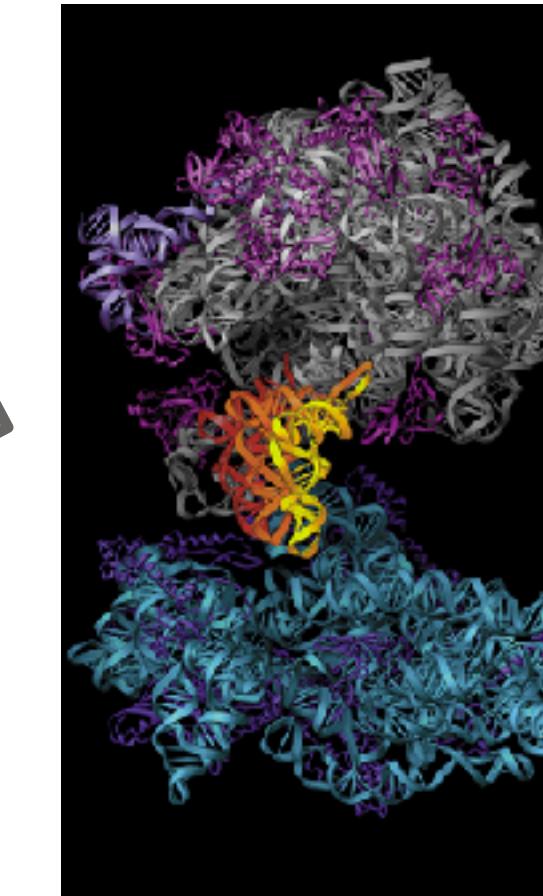
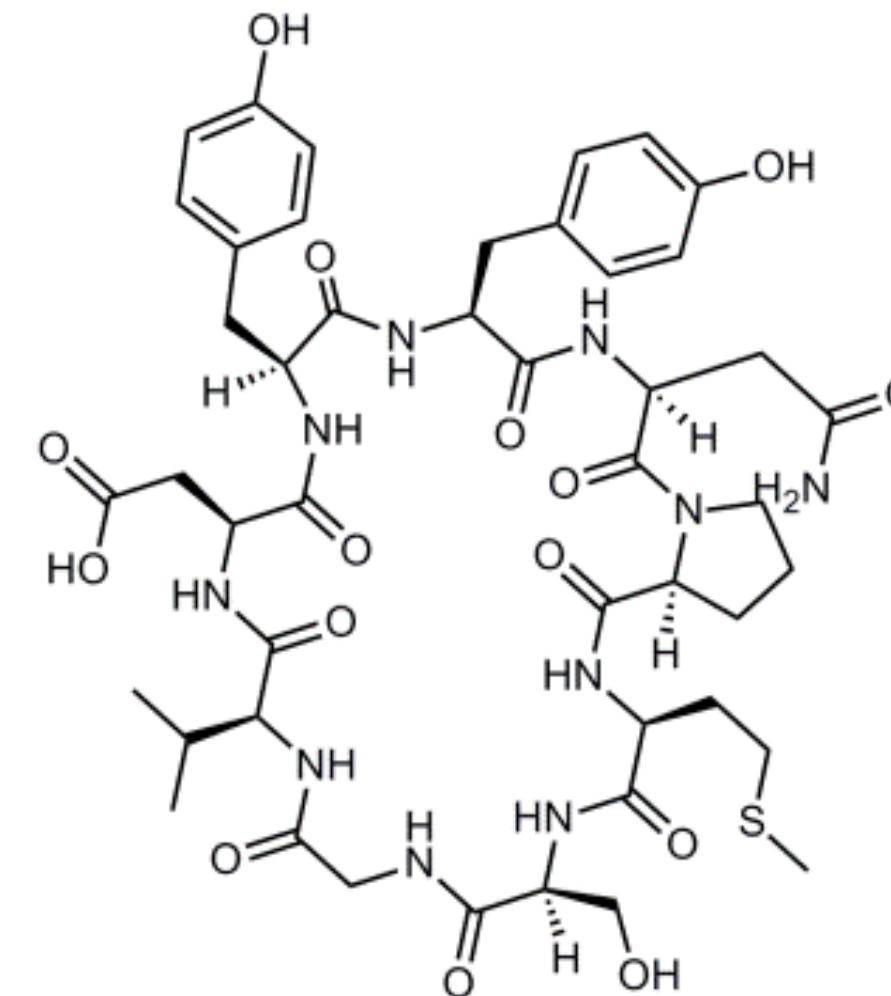
Nonribosomal biosynthetic enzymes



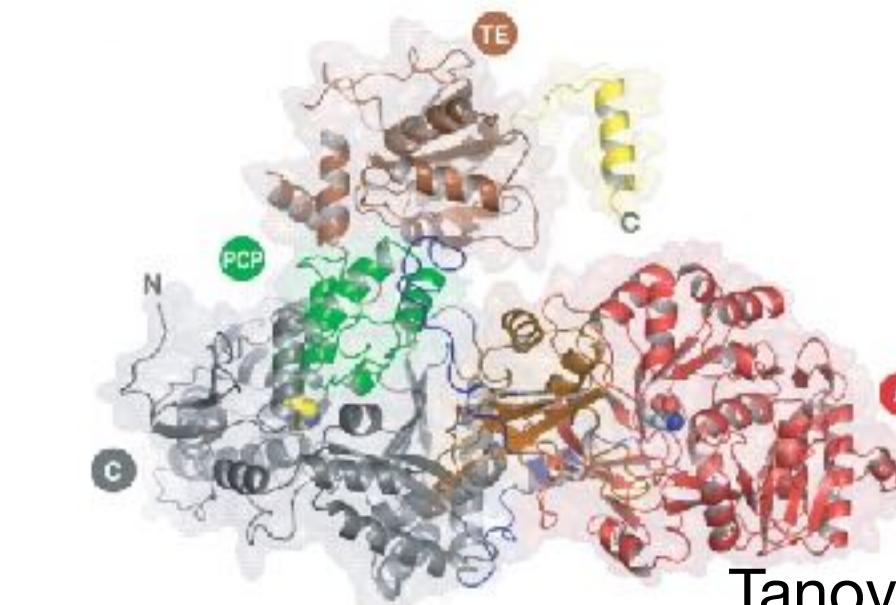
Cyanobacterial natural products



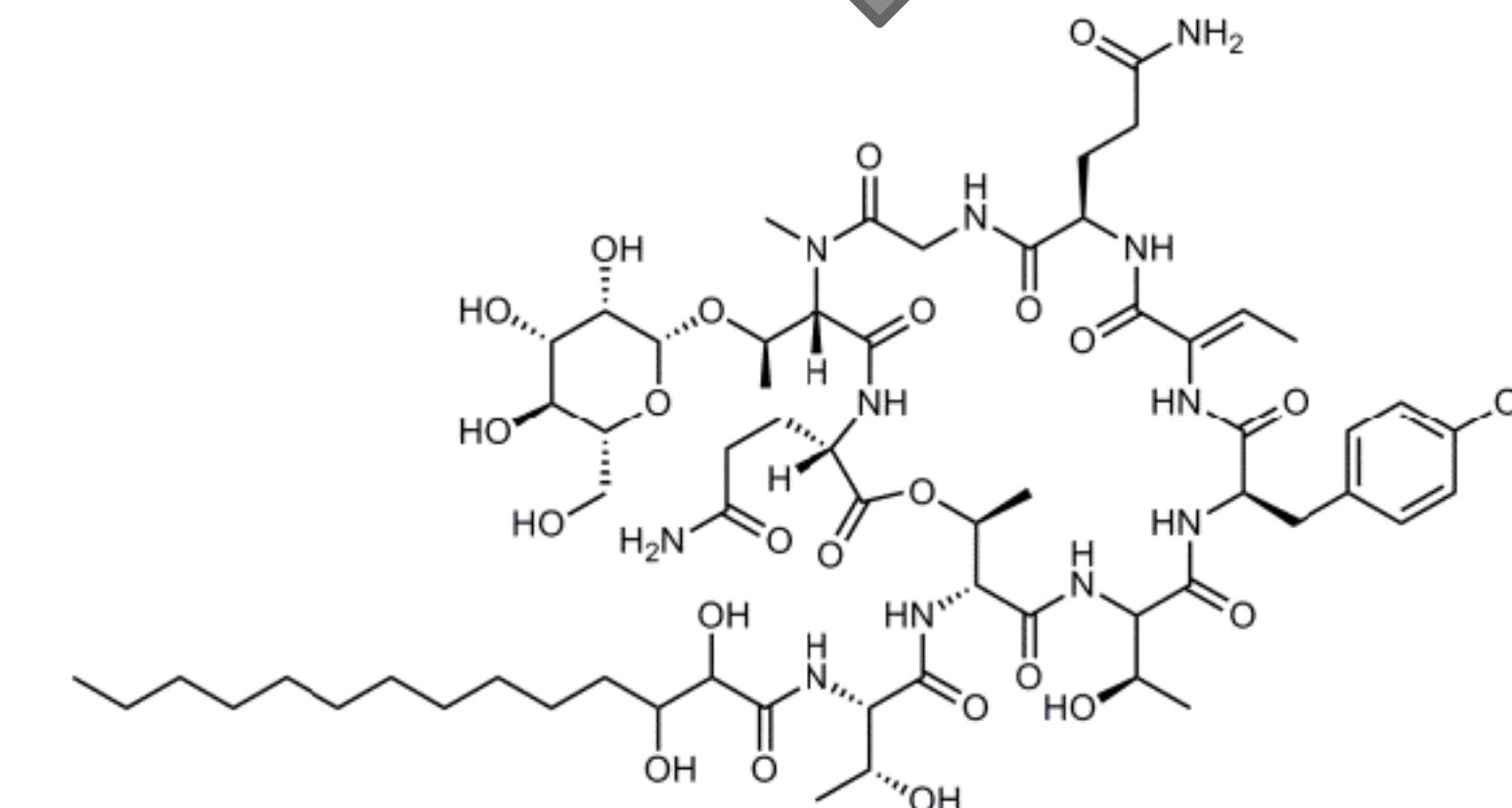
Ribosomal
pathway + post-
translational
modifications



Nonribosomal
pathway



Tanovic et al. 2008





Ribosomally synthesized and post-translationally modified (RiPP)

post-translational modifications: heterocyclization, oxidation, methylation, and prenylation

"RiPPs are produced from a short precursor peptide (PP) comprised of a leader peptide and a core peptide (Fig. 1). The PP is synthesised by the ribosome, and the core peptide is post-translationally modified by a series of RiPP tailoring enzymes (RTEs) that install various structural features onto the peptide backbone. The core peptide is usually cleaved from the leader peptide once most PTMs have been made, yielding a biologically active final product"

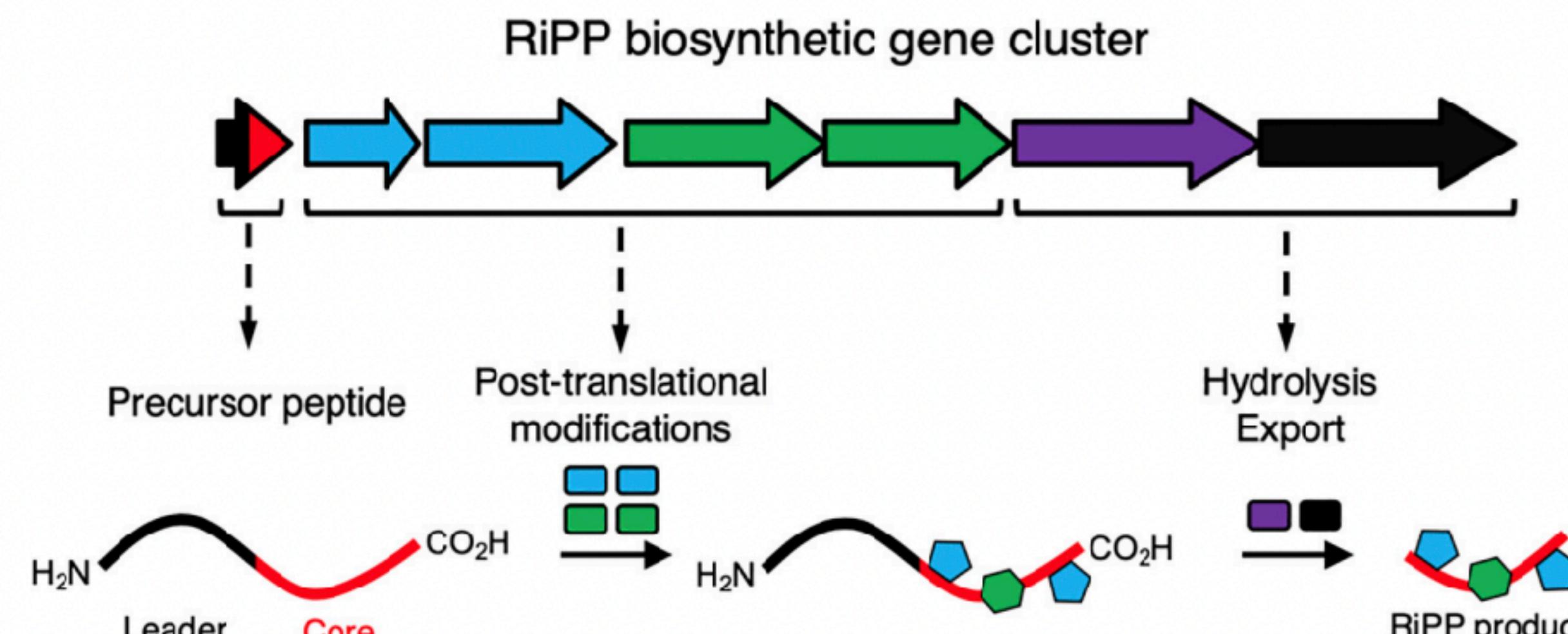


Fig. 1. Schematic of RiPP biosynthesis.



Ribosomaly synthesized and post-translationally modified (RiPP)

Table 1. Nonribosomal Peptides, Polyketides, Ribosomal Peptides, Indole Alkaloids, and Terpene Biosynthetic Gene Clusters from Cyanobacteria

Compound	Gene Clusters	Producer Strain	Compound Type	Size (kb)	Lineage ^a	Year	Refs
Aeruginosamide	CAIH01000183	<i>Microcystis aeruginosa</i> PCC 9432	Cyanobactin	12.4	I	2013	[38]
Aestuaramide	AAVU01000047	<i>Lyngbya aestuarii</i> PCC 8106	Cyanobactin	14.8	III	2014	[77]
Anacyclamide	FJ461733	<i>Anabaena</i> sp. 90	Cyanobactin	11	IV	2010	[62]
Cyanothecamide	CP001344	<i>Cyanothece</i> sp. PCC 7425	Cyanobactin	16.5	I	2012	[78]
Microcyclamide	AM931579	<i>Microcystis aeruginosa</i> PCC 7806	Cyanobactin	12.6	I	2008	[79]
Patellamide	AY986476	<i>Prochloron didemni</i>	Cyanobactin	11	P	2005	[36]
Piricyclamide	JQ951924	<i>Microcystis aeruginosa</i> PCC 7005	Cyanobactin	9.3	I	2012	[71]
Prenylagaramide	HQ655154	<i>Planktothrix agardhii</i> NIES-596	Cyanobactin	13.4	III	2011	[39]
Tenuecyclamide	EU290741	<i>Nostoc spongiaeforme</i> var. <i>tenue</i> str. Carmeli	Cyanobactin	11.8	IV	2008	[37]
Trichamide	CP000393	<i>Trichodesmium erythraeum</i> ISM101	Cyanobactin	12.5	III	2006	[66]
Trunkamide	EU290742	<i>Prochloron</i> sp. 06037A	Cyanobactin	11.1	P	2008	[37]
Viridisamide	CP003614	<i>Oscillatoria nigro-viridis</i> PCC 7112	Cyanobactin	13	III	2013	[38]
Microviridin	AM943877	<i>Microcystis aeruginosa</i> NIES-298	Microviridin	5.8	I	2008	[42]
Prochlorosin	BX548175	<i>Prochlorococcus marinus</i> MIT9313	Lantipeptide	P	I	2010	[46]



Mycosporine-like Amino Acids (MAAs) and Scytonemin Indole Alkaloids Terpenes

Compound	Gene Clusters	Producer Strain	Compound Type	Size (kb)	Lineage ^a	Year	Refs
Ambiguine	KF664586	<i>Fischerella ambigua</i> UTEX 1903	Indole-Terpene	41.7	V	2014	[52]
Welwitindolinone	KF811479	<i>Hapalosiphon welwitschii</i> UTEX B 1830	Indole-Terpene	36.6	V	2014	[53]
Welwitindolinone	KJ767018	<i>Westiella intricata</i> UH HT-29-1	Indole-Terpene	59.2	V	2014	[54]
Mycosporine-2-glycine	AB854644	<i>Aphanothecce halophytica</i>	NRPS-like	4	I	2014	[80]
Shinorine	CP000117	<i>Anabaena variabilis</i> ATCC 29413	NRPS-like	6.8	IV	2010	[48]
Cylindrocyclophane	JX477167	<i>Cylindrospermum licheniforme</i> UTEX 'B 2014'	Cyclophane	26.2	IV	2012	[81]
Scytonemin	CP001037	<i>Nostoc punctiforme</i> ATCC 29133	Indole-alkaloid	29.6	IV	2007	[49]
Saxitoxin	DQ787200	<i>Cylindrospermopsis raciborskii</i> T3	Alkaloid	37.6	IV	2008	[64]
Geosmin	CP001037	<i>Nostoc punctiforme</i> PCC 73102	Terpene	2.3	IV	2009	[58]
2-Methylisoborneol	HQ830028	<i>Pseudanabaena</i> sp. dqh15	Terpene	5.9	III	2011	[59]



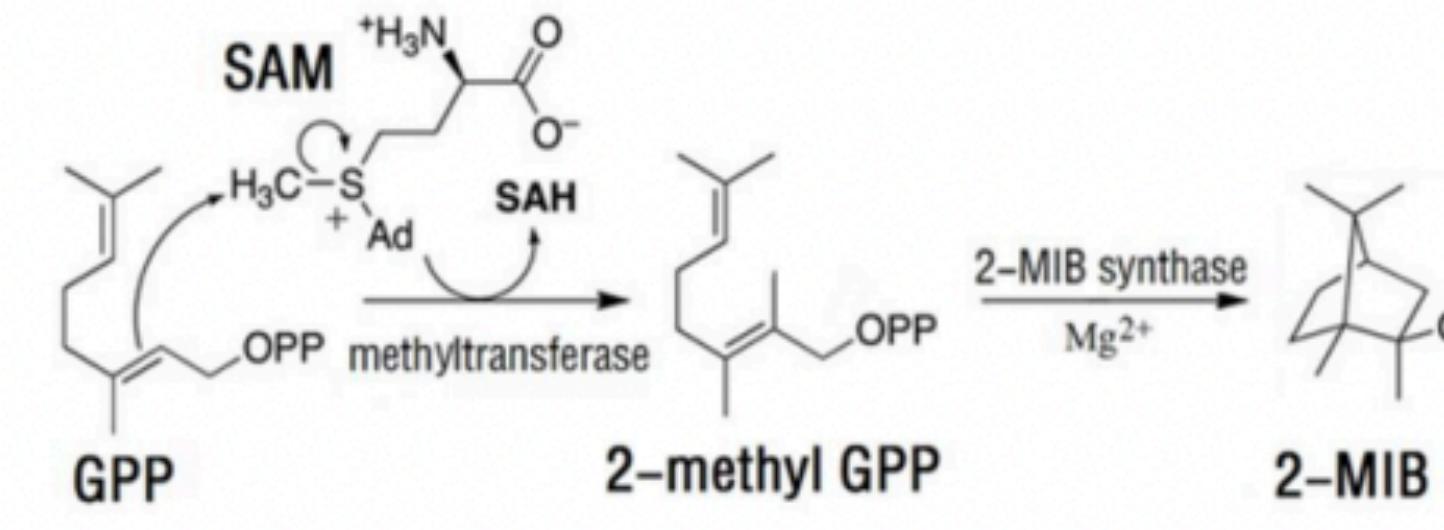
Terpenes

medicines, flavors, fragrances, pigments, and insecticides

Geosmin - earthy smell

2-Methylisoborneol - musty or earthy odor

A) Geranyl pyrophosphate (GPP)



2-methyltransferase 2-MIB synthase nucleotide-binding protein



B) Farnesyl diphosphate (FPP)

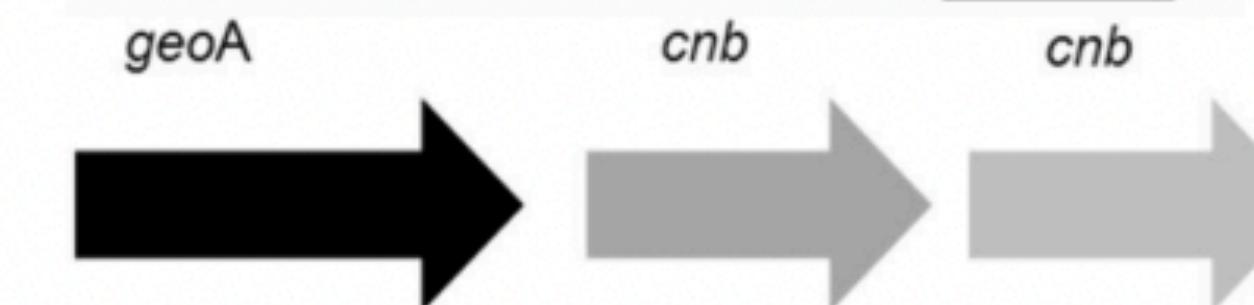
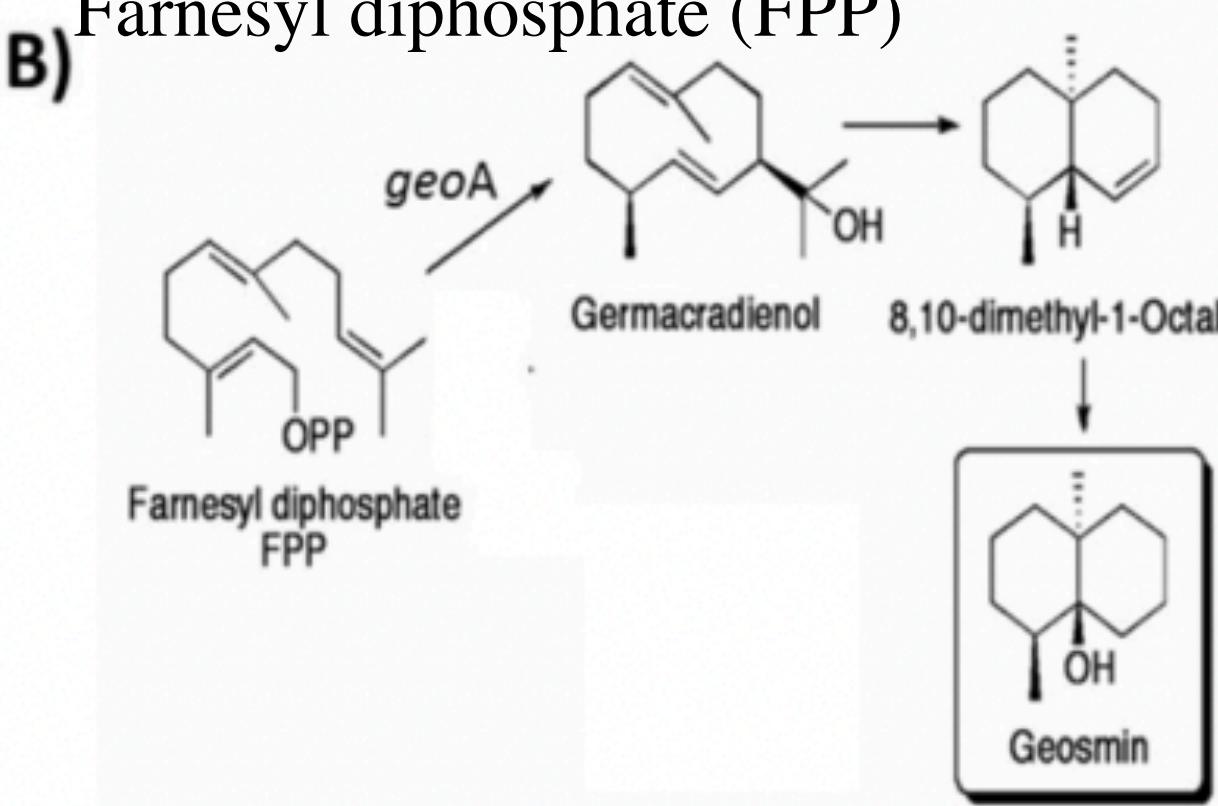


Figure 4. Biosynthetic gene clusters and biosynthetic mechanism of A) 2-methylisoborneol and B) geosmin. GPP=geranyl pyrophosphate and cnb=cyclic nucleotide binding protein gene. (Figure is modified from Giglio et al. 2008 and Wang et al. 2011.)



Example of study

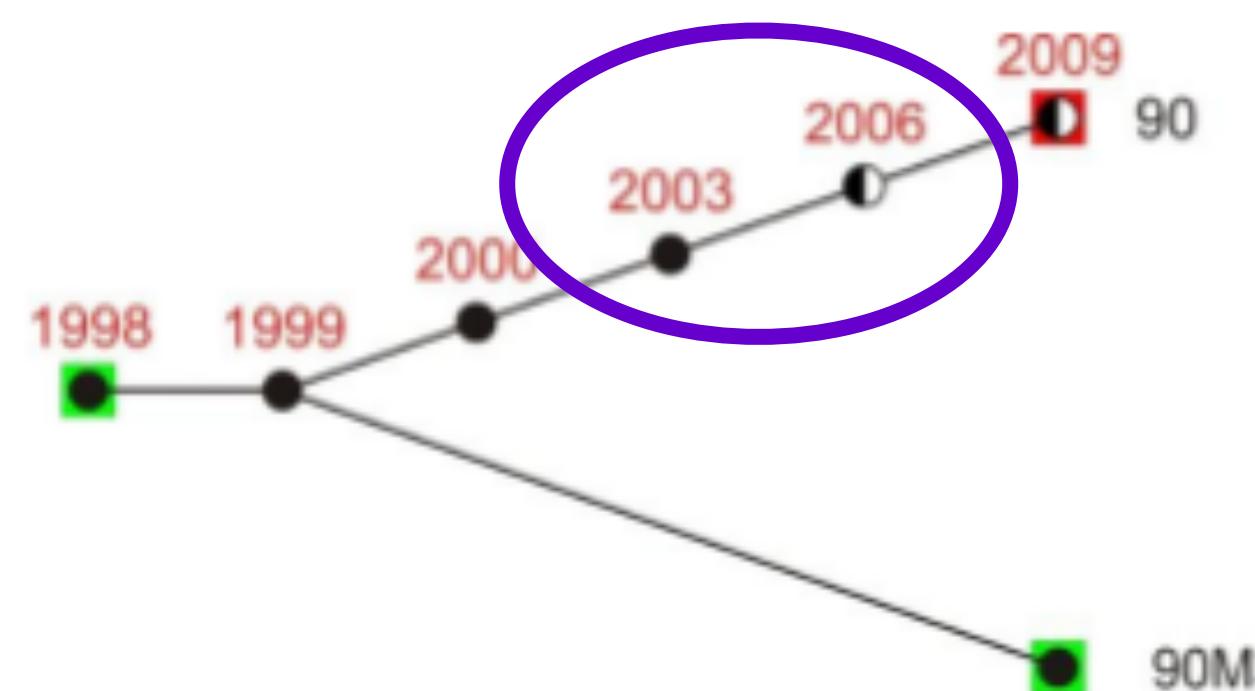
Hassallidins, antifungal glycolipopeptides, are widespread among cyanobacteria and are the end-product of a nonribosomal pathway

Anabaena sp. 90

Hassallidin (ANA_C13055-C13074, 62 kb)



Wang et al. 2012 BMC Genomics

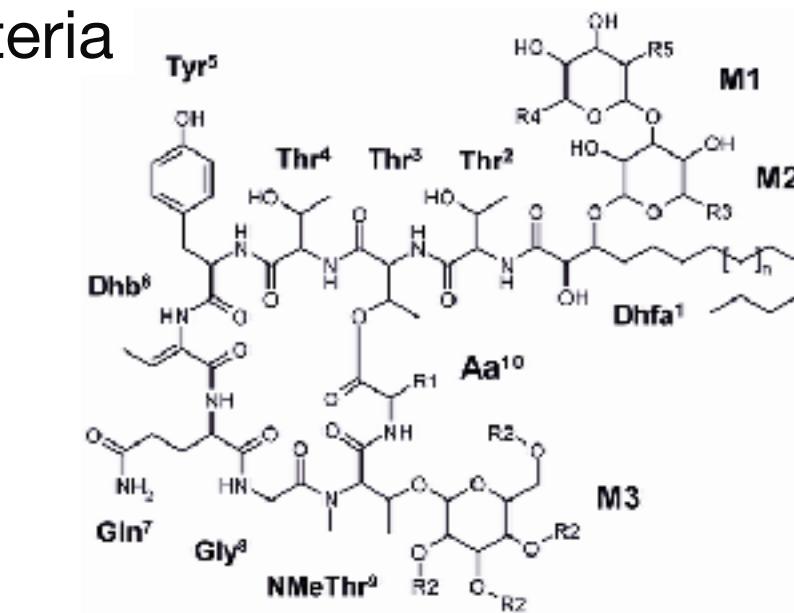


Vestola et al. 2014 PNAS

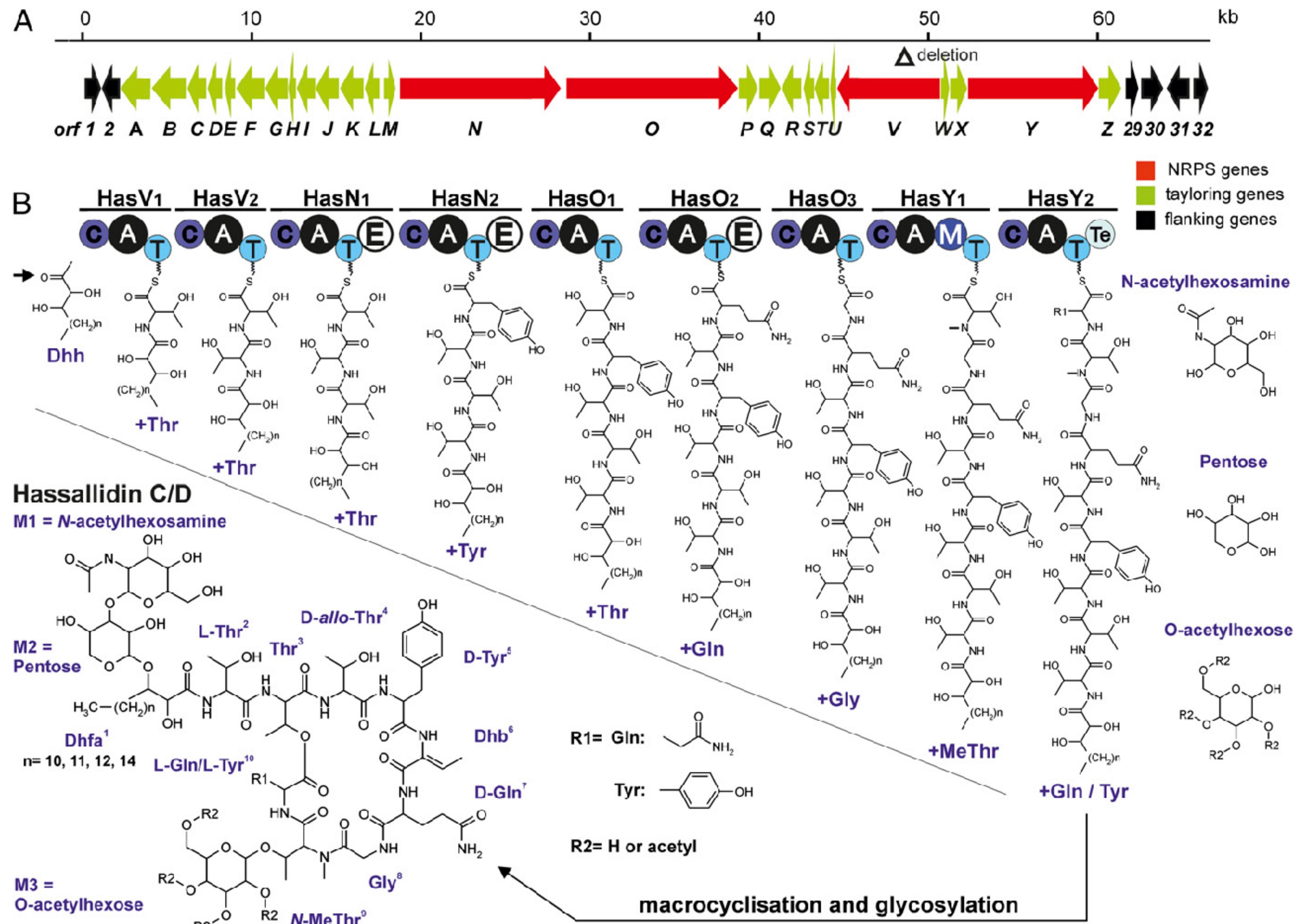
Table 2. Detected (LC/MS) hassallidin production of cyanobacterial strains and results of PCR amplification of *has* NRPS genes indicating the presence of the *has* gene cluster

No.	Strain code	Origin, location	Year	has NRPS genes	Hassallidins LC/MS
	<i>Anabaena</i> sp.				
1	90, y 2009	Lake Vesijärvi, Finland	1986	■	□
2	90, y 1998 (freeze dried)	Lake Vesijärvi, Finland	1986	■	■
3	90 M3 (apdA mutant)	Lake Vesijärvi, Finland	1986	■	■
4	299A	Lake Vesijärvi, Finland	1992	■	□
5	299B	Lake Vesijärvi, Finland	1992	■	■
6	258	Lake Tuusulanjärvi, Finland	1990	■	■
7	SYKE748A	Lake Tuusulanjärvi, Finland	1999	■	■
8	SYKE763A	Lake Tuusulanjärvi, Finland	1999	■	■
9	OTU33S 16	Lake Tuusulanjärvi, Finland	2000	■	■
10	OTU43S8	Lake Tuusulanjärvi, Finland	2000	■	■
11	1TU33S8	Lake Tuusulanjärvi, Finland	2001	■	■
12	1TU35S 12	Lake Tuusulanjärvi, Finland	2001	■	■
13	1TU44S9	Lake Tuusulanjärvi, Finland	2001	■	■
14	1TU44S 16	Lake Tuusulanjärvi, Finland	2001	■	■
15	SYKE971/6	Lake Kotojärvi, Finland	1999	■	■
16	PH256	Lake Knud, Denmark	1994	■	□
17	NIVA-CYA 269/2	Lake Frøylandsvatnet, Norway	1990	■	■
18	NIVA-CYA 269/6	Lake Frøylandsvatnet, Norway	1990	■	■
19	XSPORK5C	Porkkala Cape, the Baltic Sea coast, Finland	1999	■	■
20	XSPORK7B	Porkkala Cape, the Baltic Sea coast, Finland	1999	■	■
21	XSPORK36B	Porkkala Cape, the Baltic Sea coast, Finland	1999	■	■
22	XSPORK14D	Porkkala Cape, the Baltic Sea coast, Finland	1999	■	■
23	BECID19	The Gulf of Finland, Vuosaari, Finland	2001	■	■
	<i>Cylindrospermopsis raciborskii</i>				
24	ATC-9502	Lake Balaton, Hungary	1994	■	■
25	CS-505	Freshwater, Solomon Dam, Australia	1996	■	■
	<i>Aphanizomenon gracile</i>				
26	Heaney/Camb 1986 140 1/1	Freshwater, Lough Neagh, Ireland	1986	■	■
	<i>Nostoc</i> sp.				
27	159	Lake Haukkajärvi, Finland	1986	■	■
28	113.5	Lichen associated	?	■	■
	<i>Tolyphothrix</i> sp.				
29	PCC 9009	Watkins Glen State Park, New York, United States	?	■	■
30	PCC 7504	Aquarium, Stockholm, Sweden	1972	■	□
31	PCC 7101	Borneo, soil	1950	■	□

Filled square, positive result; unfilled square, negative result. See also SI Appendix, Table S5.



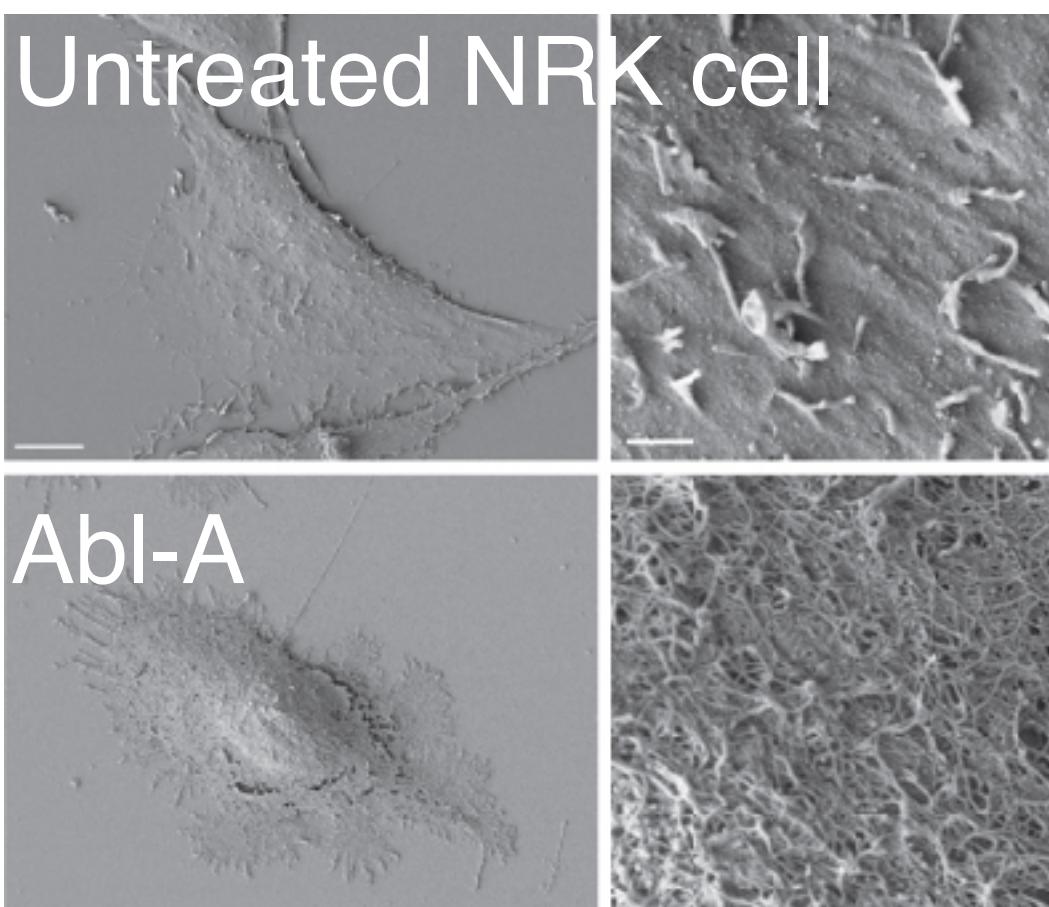
Example of study



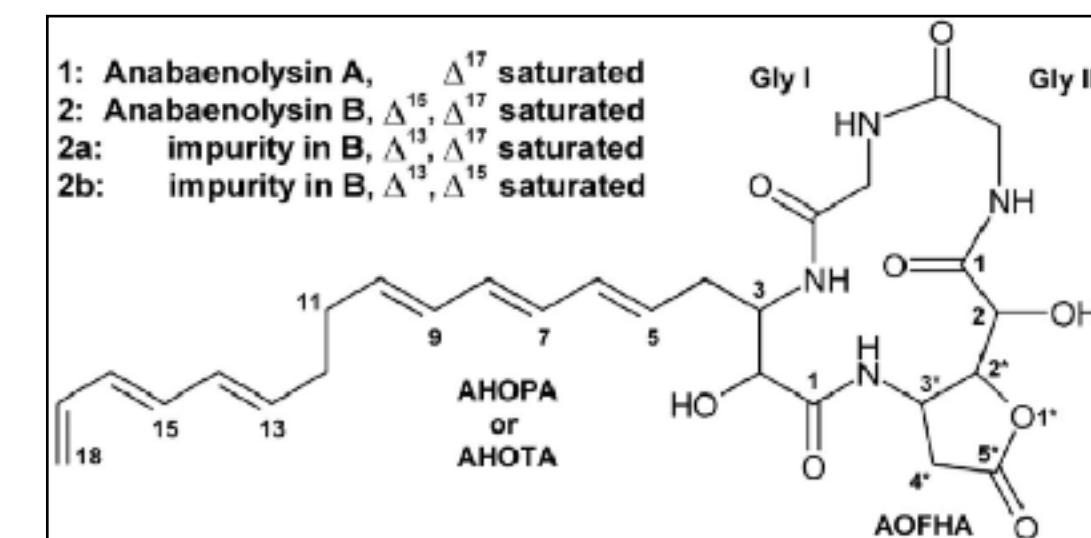
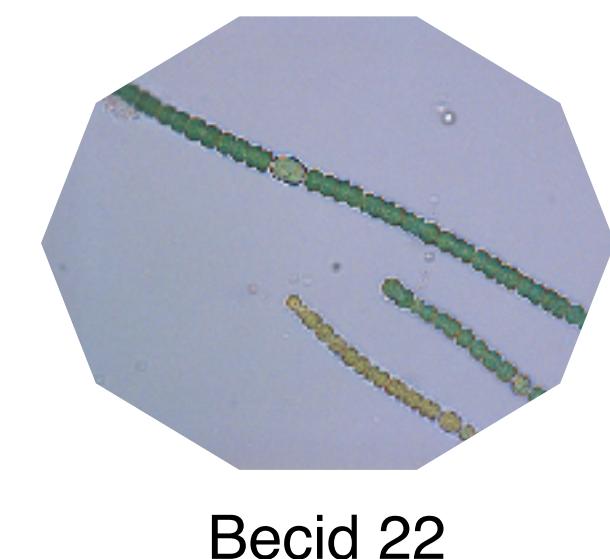


Example of study

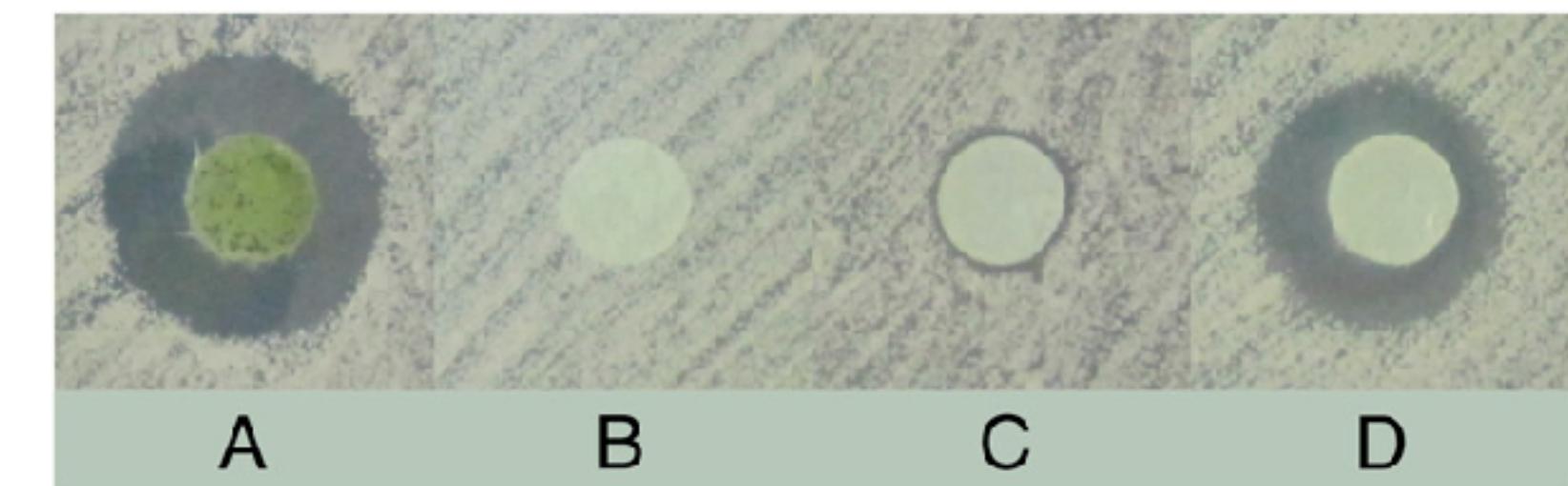
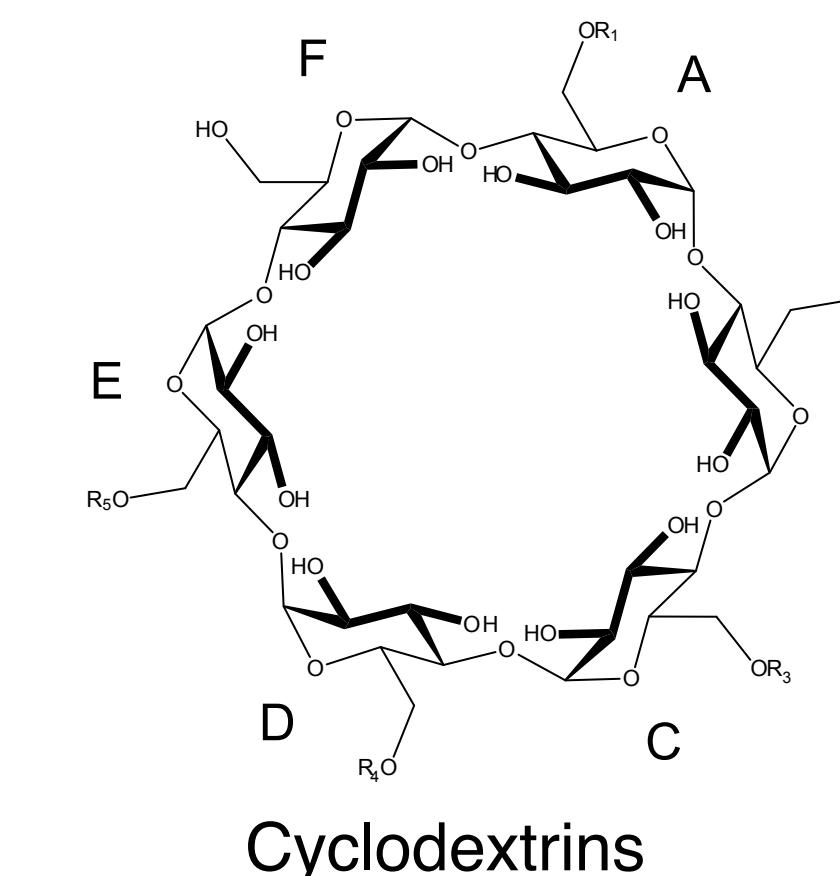
Antifungal activity improved by coproduction of cyclodextrins and anabaenolysins in Cyanobacteria



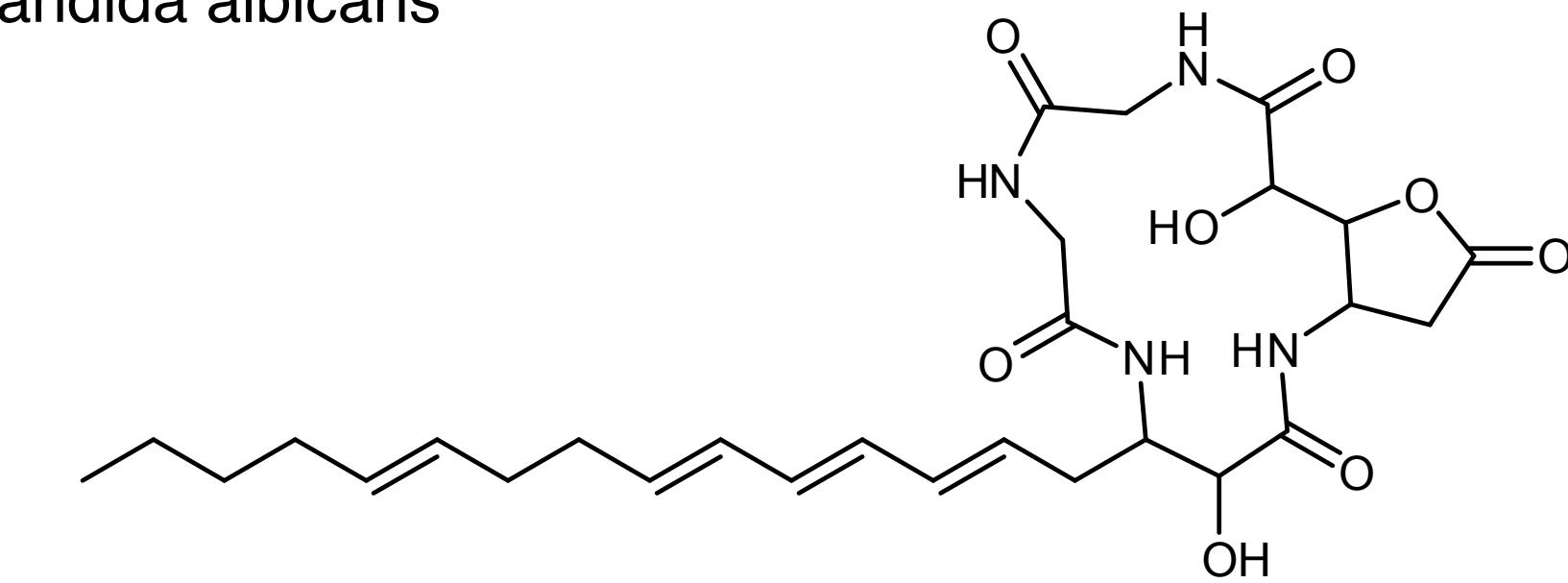
Oftedal et al. 2012 Biochem Biophys Acta



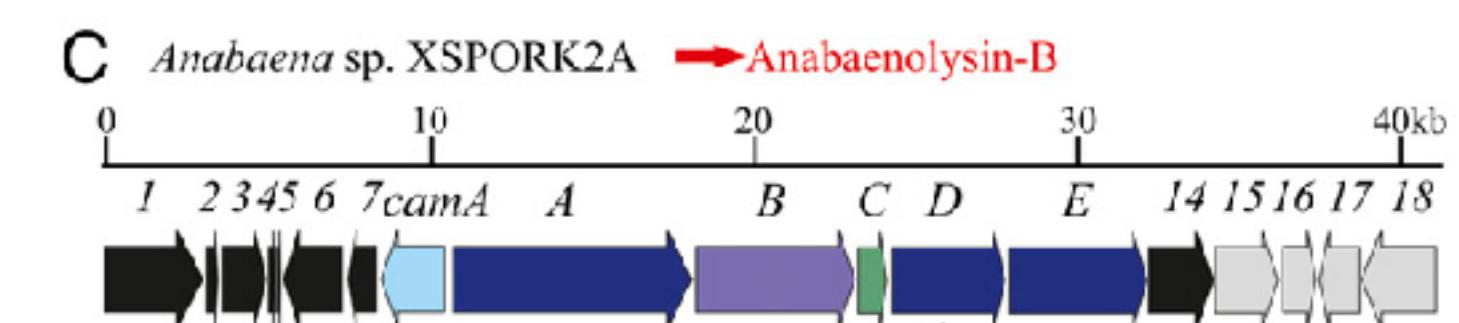
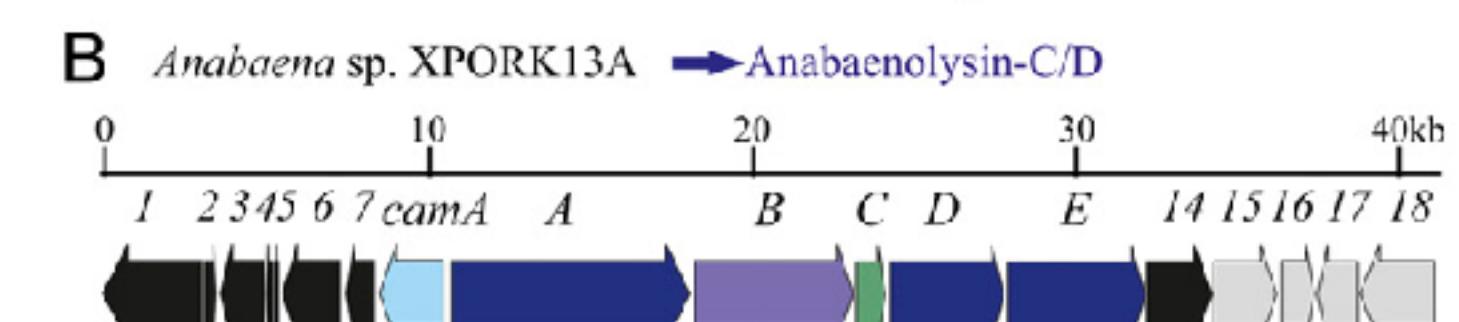
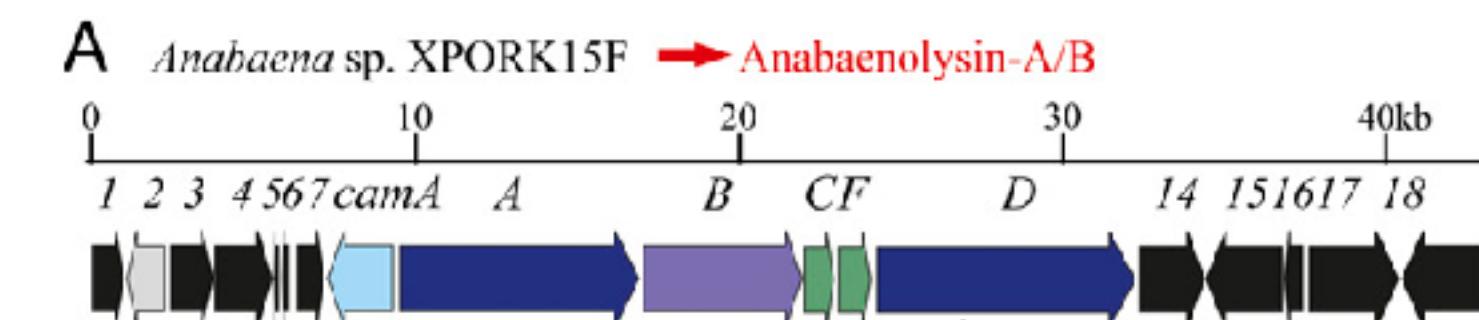
Jokela et al. 2012 PlosOne



Candida albicans



Anabaenolysin B

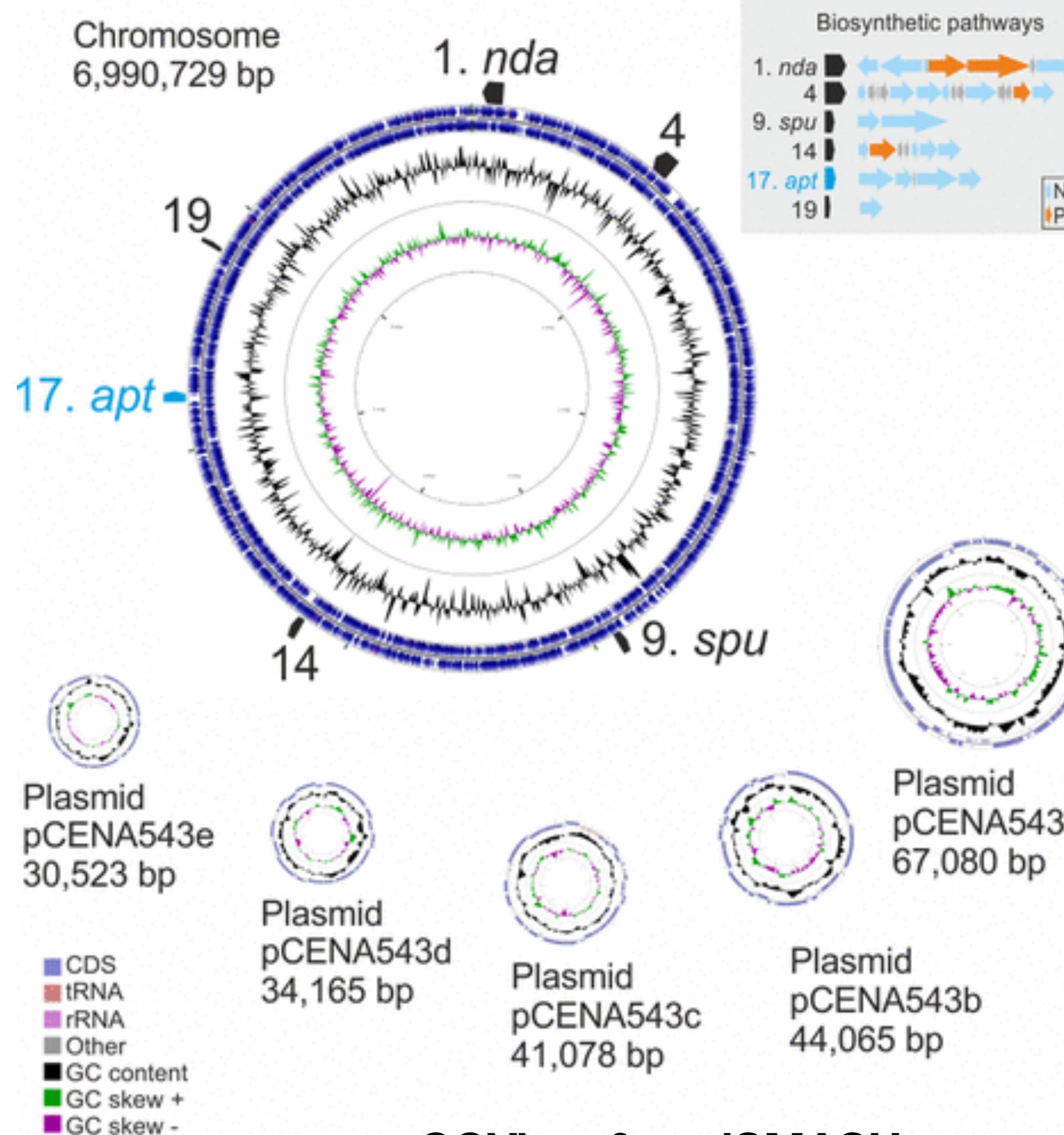


Shishido et al. 2015 ACS Chem Biol



Example of study

Simultaneous Production of Anabaenopeptins and Namalides by the Cyanobacterium *Nostoc* sp. CENA543



CGView & antiSMASH

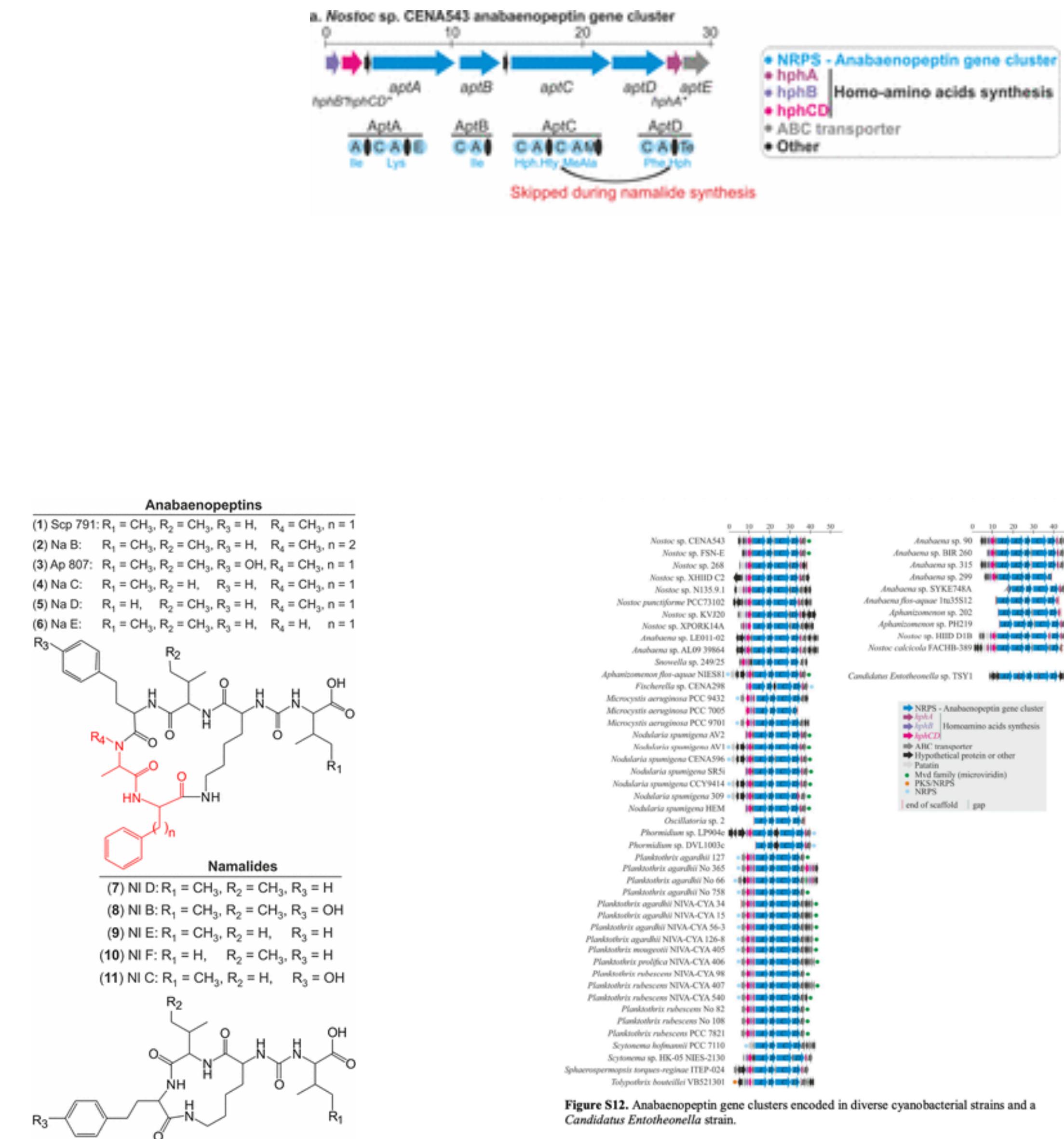
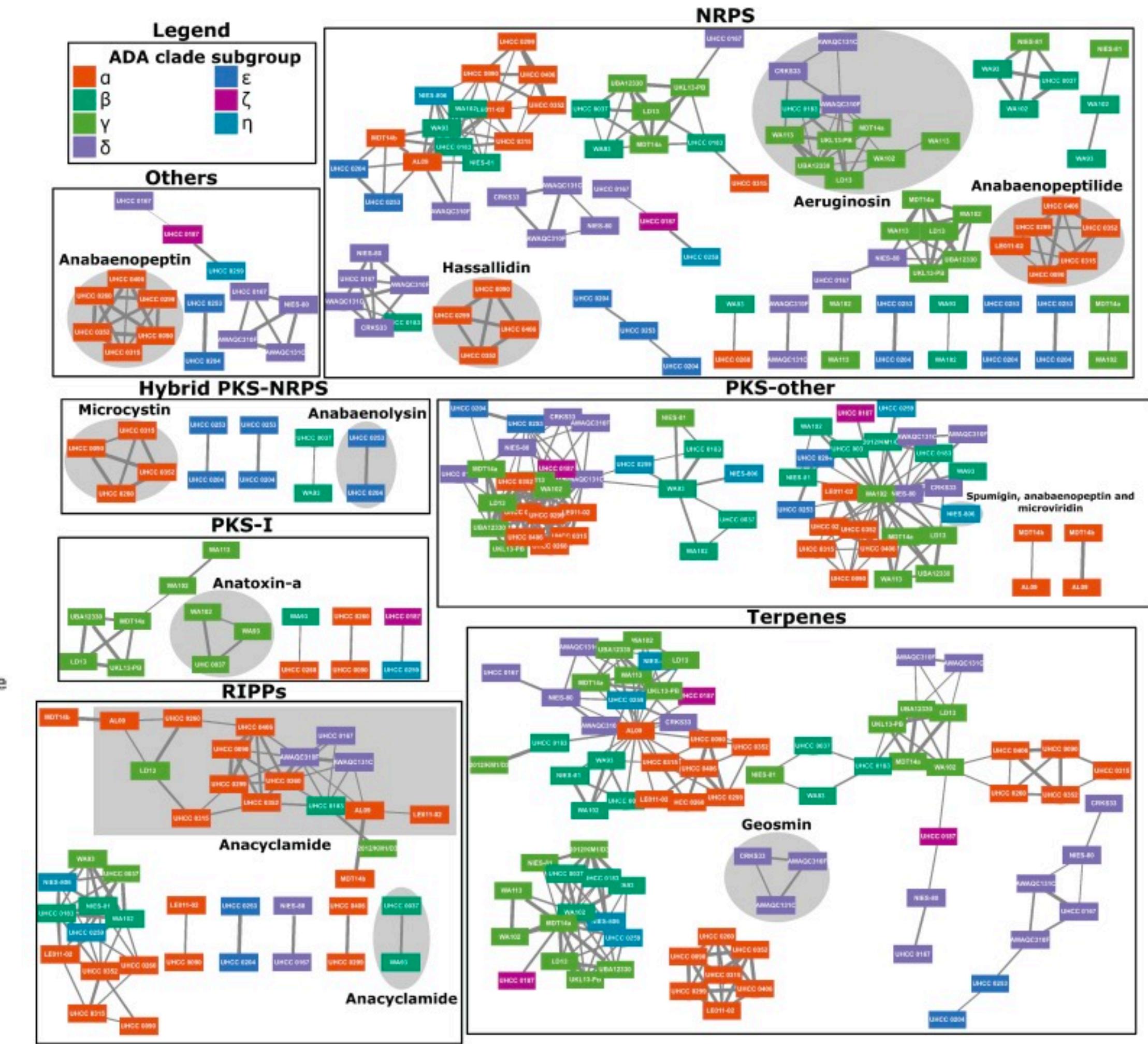
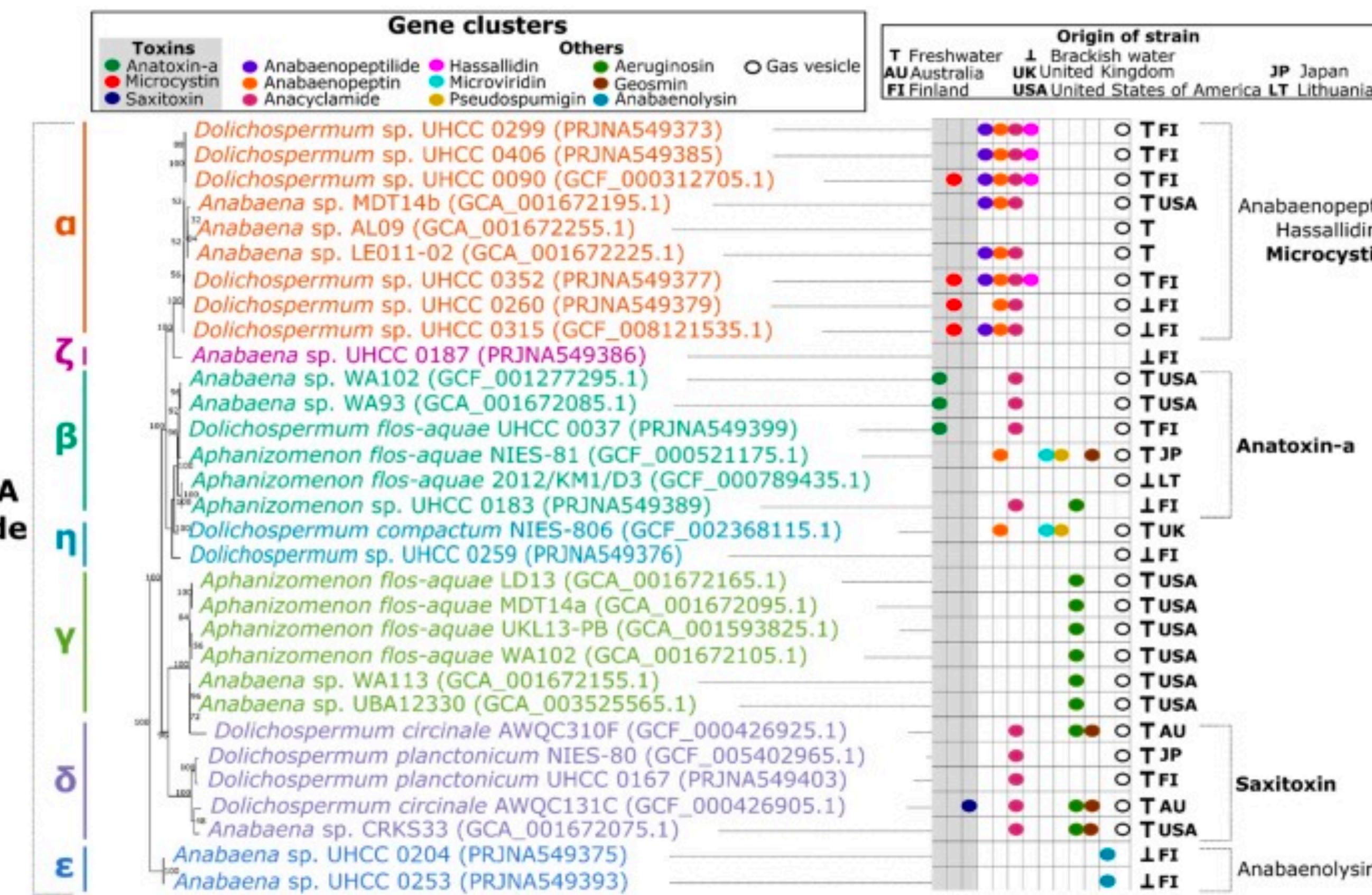


Figure S12. Anabaenopeptin gene clusters encoded in diverse cyanobacterial strains and a *Candidatus Entotheonella* strain.



Example of study

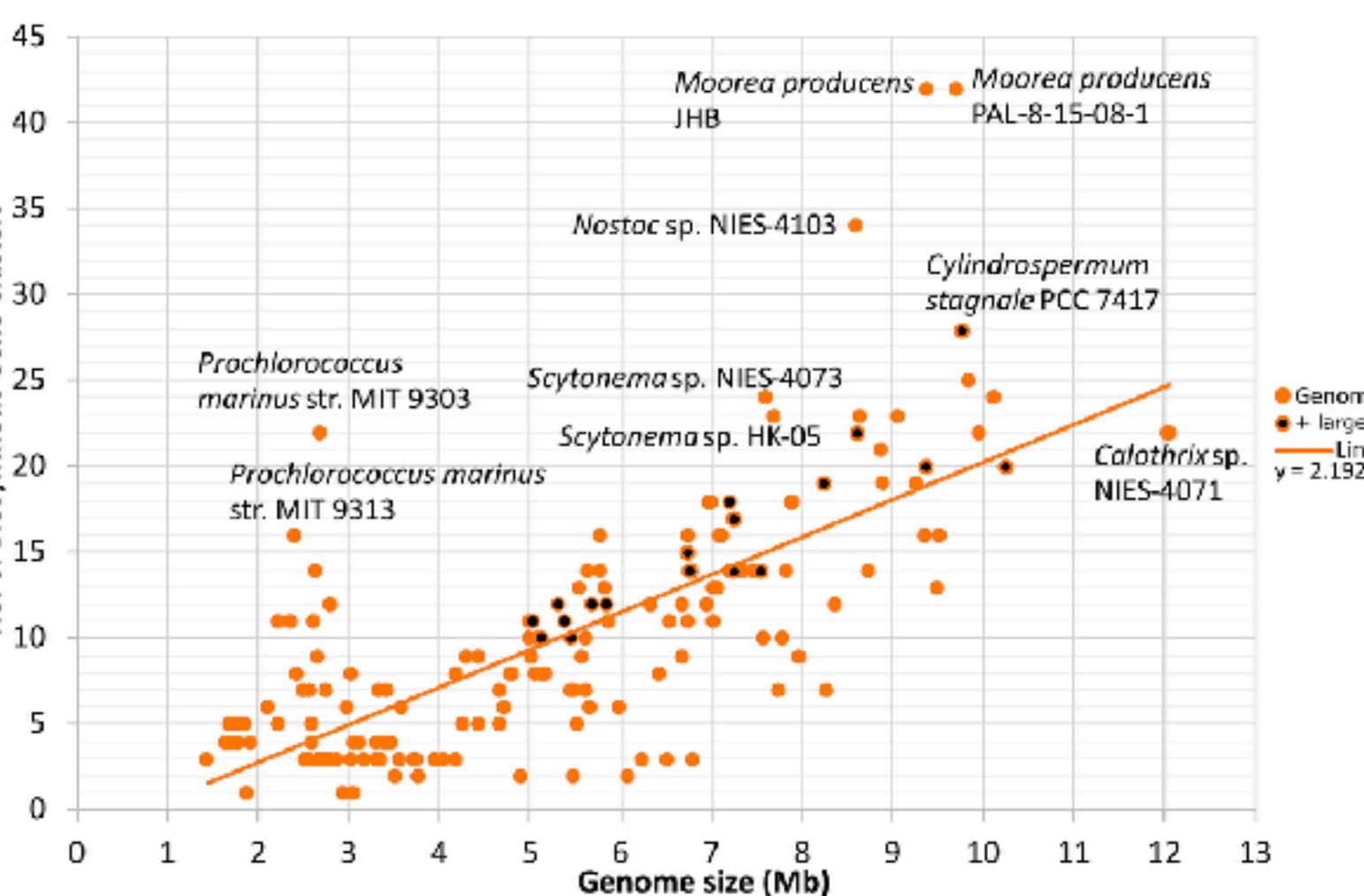
Phylogenomic Analysis of Secondary Metabolism in the Toxic Cyanobacterial Genera *Anabaena*, *Dolichospermum* and *Aphanizomenon*





Example of study

Mining of Cyanobacterial Genomes Indicates Natural Product Biosynthetic Gene Clusters Located in Conjugative Plasmids

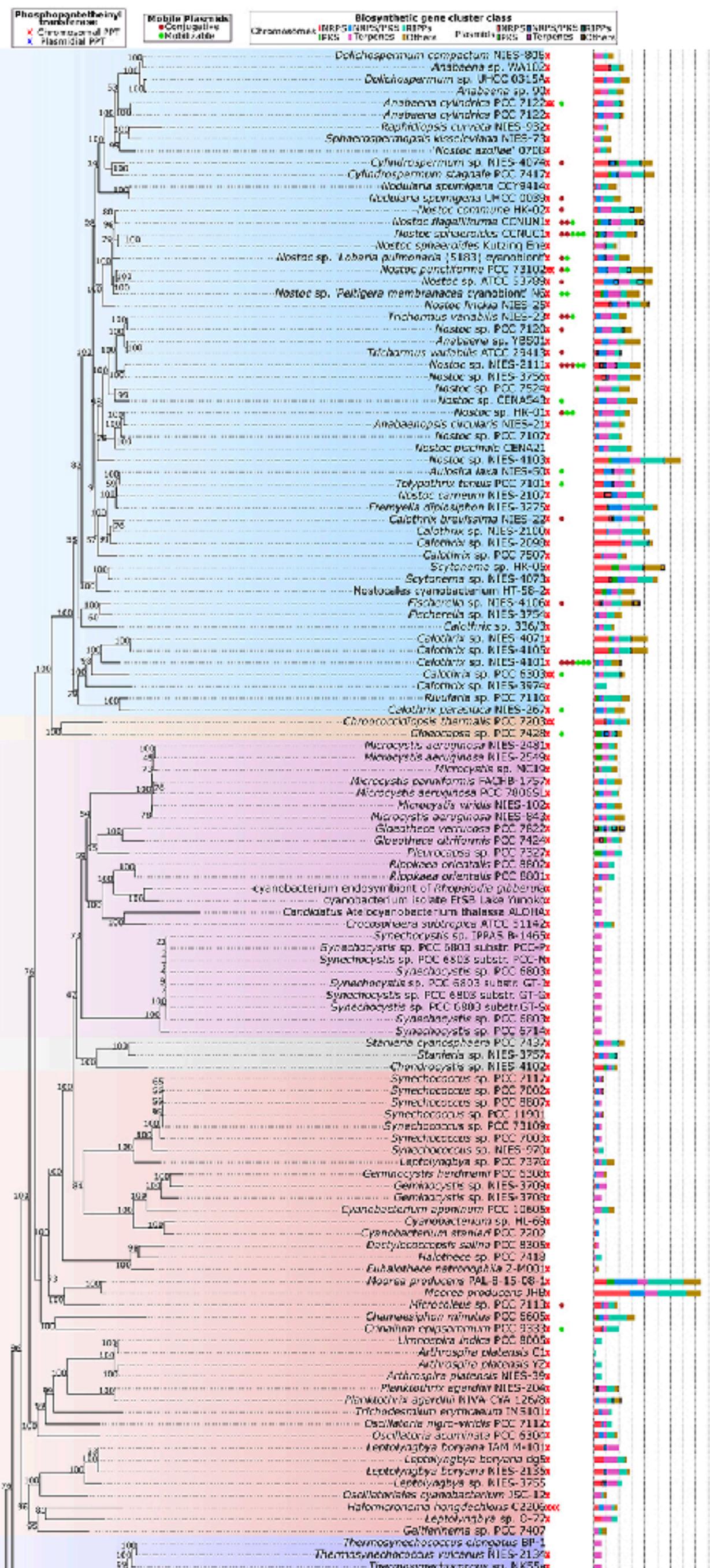


185 analyzed genomes revealed 1817 natural products BGCs.

Individual genomes contained 1–42 biosynthetic pathways (mean 8)

95% of which were present in chromosomes and the remaining 5% in plasmids.

Popin et al 2021 Front Microbiol





BGC databases

Compounds

Cluster mining tools

2metDB

antiSMASH

ARTS

BAGEL

BiG-SCAPE

CASSIS and SMIPS

CLUSEAN

ClusterFinder

ClustScan Professional

eSNaPD // environmental Surveyor
of Natural Product Diversity

EvoMining

FunGeneClusterS

MIDDAS-M

MIPS-CG

NaPDoS // Natural Products Domain
Seeker

PhytoClust

PKMiner

plantiSMASH

PRISM / GNP

RiPPMiner

RODEO

SANDPUMA

SBSPKS

SeMPI

Docs » Cluster mining tools

Tools for mining secondary metabolite biosynthesis gene clusters

Table of Contents

- Tools for mining secondary metabolite biosynthesis gene clusters
 - 2metDB
 - antiSMASH
 - ARTS
 - BAGEL
 - BiG-SCAPE
 - CASSIS and SMIPS
 - CLUSEAN
 - ClusterFinder
 - ClustScan Professional
 - eSNaPD // environmental Surveyor of Natural Product Diversity
 - EvoMining
 - FunGeneClusterS
 - MIDDAS-M
 - MIPS-CG
 - NaPDoS // Natural Products Domain Seeker
 - PhytoClust
 - PKMiner
 - plantiSMASH
 - PRISM / GNP
 - RiPPMiner
 - RODEO
 - SANDPUMA
 - SBSPKS
 - SeMPI
 - SMURF / Secondary Metabolite Unknown Region Finder

<https://www.secondarymetabolites.org/mining/>



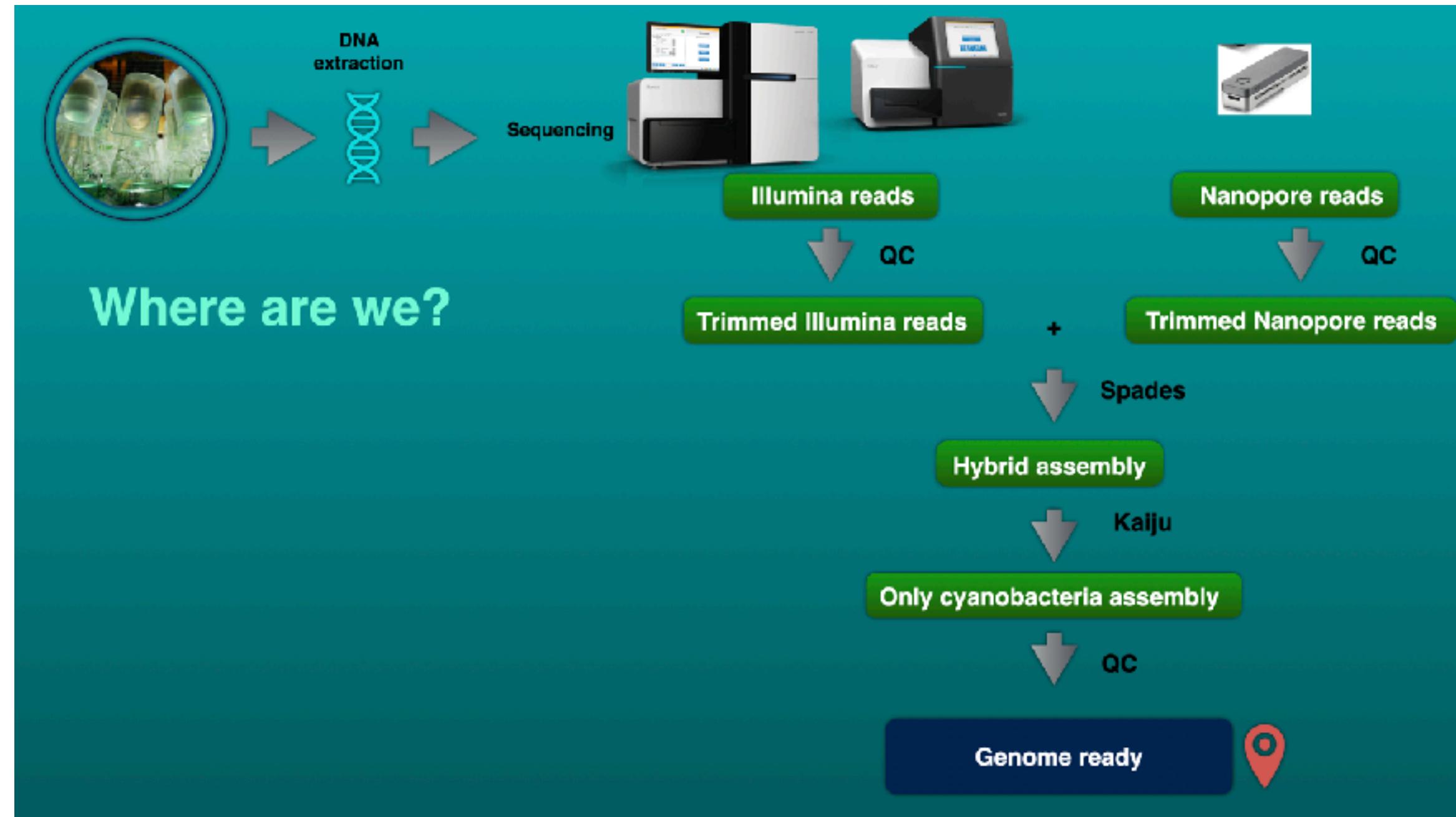
Ribosomaly synthesized and post-translationally modified (RiPP)

Table 1
Summary of genome mining tools available for RiPPs.

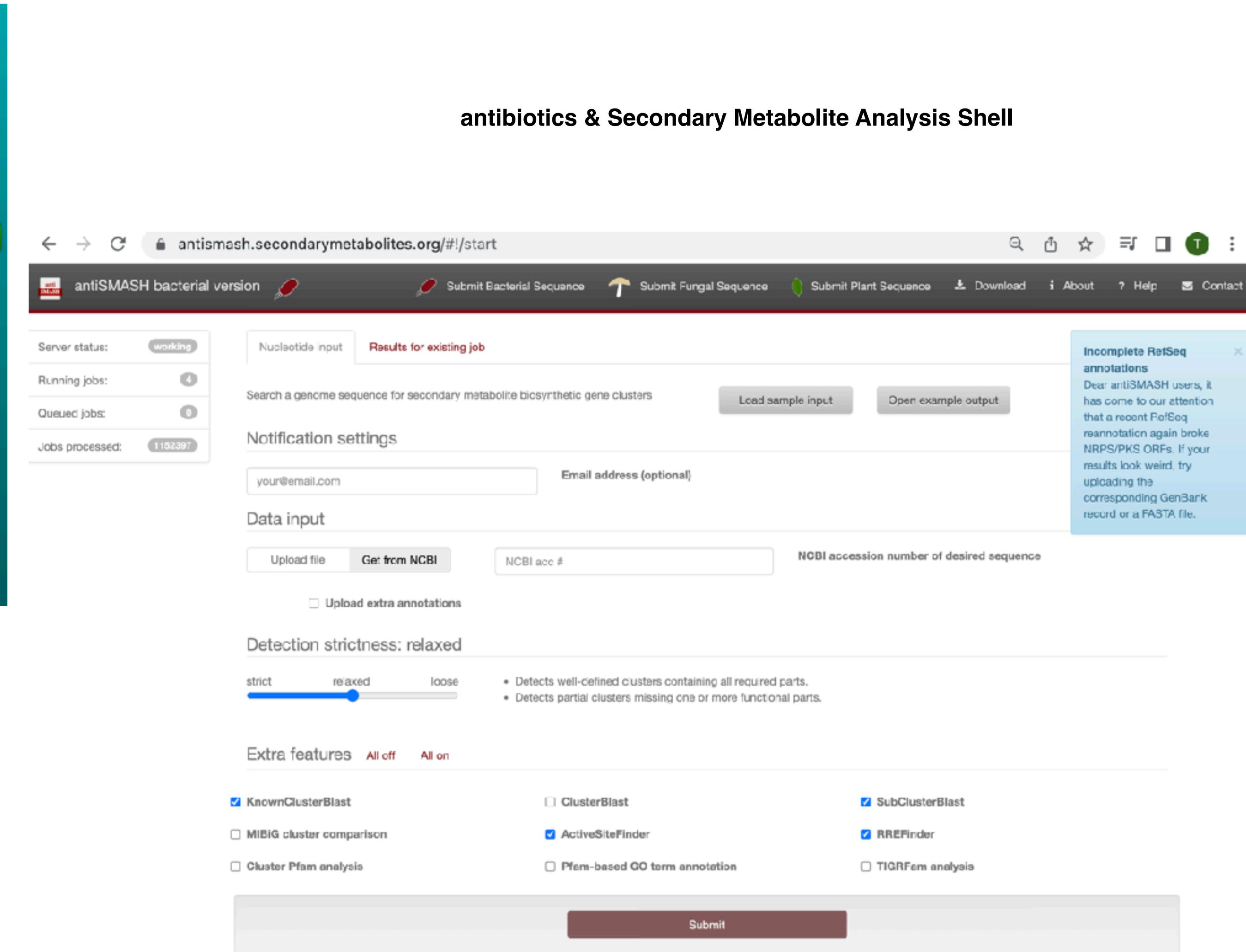
Tool	Web address	Function and RiPP class	Interface	Input	Output
BAGEL4	http://bagel4.molgenrug.nl/	<ul style="list-style-type: none"> • BGC identification and annotation Web • Multiple RiPP classes 	Web	Sequence file (FASTA) or built-in set of publicly available genomes in RefSeq database	<ul style="list-style-type: none"> • Html output showing BGC regions with gene annotations • Sequence alignment with curated precursor peptides • Downloadable GenBank files, FASTA files, gene tables and promoter/terminator information
antiSMASH5	https://antismash.secondarymetabolites.org/	<ul style="list-style-type: none"> • BGC identification, annotation and analysis • Multiple RiPP classes 	Web	Sequence file (FASTA, GenBank or EMBL) or NCBI nucleotide accession	<ul style="list-style-type: none"> • Html output showing BGC regions with gene annotations and predicted class • Predicted PP and cleavage sites for some RiPP classes • Downloadable GenBank files and other data for BGC regions • KnownClusterBlast analysis
PRISM4	http://grid.adapsyn.com/prism/#!/prism	<ul style="list-style-type: none"> • BGC identification, PP cleavage and PTM prediction • Multiple RiPP classes 	Web	Sequence file (FASTA or GenBank)	<ul style="list-style-type: none"> • Html output showing BGC regions with gene annotations and predicted class • Predictions of core peptide and final structures • SMILES strings for predicted structures, FASTA sequences of BGCs
RiPPMiner	http://202.54.226.242/~priyesh/rippminer2/new_predictions/index.php	<ul style="list-style-type: none"> • BGC identification and RiPP class • Predictions of structure, cleavage and crosslinks • Multiple RiPP classes 	Web	Peptide = PP sequence (raw or FASTA) Genome = sequence file (FASTA)	<p>Peptide</p> <ul style="list-style-type: none"> • Html output with predicted structure and class • SMILES strings for predicted structures <p>Genome</p> <ul style="list-style-type: none"> • Html output showing identified clusters and annotations as well as peptide cleavage, crosslinks and structural predictions • SMILES strings of predicted structures • List of other small ORFs present in BGC • Html files with BGC information and Pfam domain annotation • .csv files of PP sequences and BGC Pfam domains
RODEO2	http://ripp.rodeo/index.html	<ul style="list-style-type: none"> • RiPP BGC identification, PP identification and structural prediction • Lasso peptides, lanthipeptides, thiopeptides & sactipeptides 	Web or Python	List of bait protein accession numbers. Optional: HMMs and configuration file	
RiPPER	https://github.com/streptomyces/ripper	<ul style="list-style-type: none"> • PP and BGC recognition • Class independent 	Docker	List of bait protein accession numbers	<ul style="list-style-type: none"> • GenBank files of retrieved BGCs annotated with short peptides • Table of PP data • RODEO files for retrieved BGCs • File of sequences classified by NeuRiPP as positive PPs • Separate file of non-RiPP peptides
NeuRiPP	https://github.com/emzodls/neuripp	<ul style="list-style-type: none"> • PP recognition • Class independent 	Python	PP sequence file (FASTA)	

Table 2
Summary of MS-based mining tools available for RiPPs.

Tool	Web address	Function and RiPP class	Interface	Input	Output
RiPPquest/ MetaMiner	https://github.com/ablab/npdtools http://gnps.ucsd.edu/ProteoSAFe/static/gnps-theoretical.jsp	<ul style="list-style-type: none"> • MS-guided genome mining, optimised for large datasets • Multiple RiPP classes 	Python or web (GNPS)	LC-MS/MS data file (MGF, mzXML, mxML or mzData) and sequence file (FASTA, antiSMASH GenBank output or BOA txt output)	<ul style="list-style-type: none"> • .tsv files with information about identified peptides and RiPP class
Pep2Path	http://pep2path.sourceforge.net/	• BGC identification from peptide MS data	Python	Comma-separated sequence of mass shifts or amino acids, and a sequence file (FASTA, GenBank or EMBL)	<ul style="list-style-type: none"> • Table with best peptide matches
CycloNovo	https://github.com/bbehsaz/cyclonovo https://gnps.ucsd.edu/ProteoSAFe/index.jsp?params=%7B%22workflow%22%22CYCLONOVO%22%7D	• Cyclopeptide identification and prediction	Python or web (GNPS)	MS data file (mzXML or MGF)	<ul style="list-style-type: none"> • MGF file of identified cyclopeptide spectra • Spectra listed with cyclopeptide scoring (txt) • Peptide sequencing reconstructions (txt)
DeepRiPP	http://deepripp.magarveylab.ca/	<ul style="list-style-type: none"> • PP structural and class predictions • BGC identification • Multiple RiPP classes 	Web	NLPrecursor: PP sequence (FASTA) BARLEY: core peptide sequence and RTE CLAMS: MS data (mzML) DeepRiPP (full): sequence file (FASTA) and optional MS file (mzML)	<ul style="list-style-type: none"> • NLPrecursor: Html output of predicted RiPP class and cleavage site • BARLEY: Html output of alignment with similar RiPPs and structure predictions • CLAMS: Html output with list of MS peaks • DeepRiPP (full): integrated Html output of NLPrecursor, BARLEY and CLAMS • Attempted matching between structure prediction and MS data



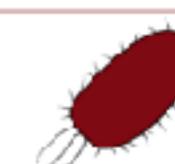
Identification of secondary metabolites biosynthetic gene clusters

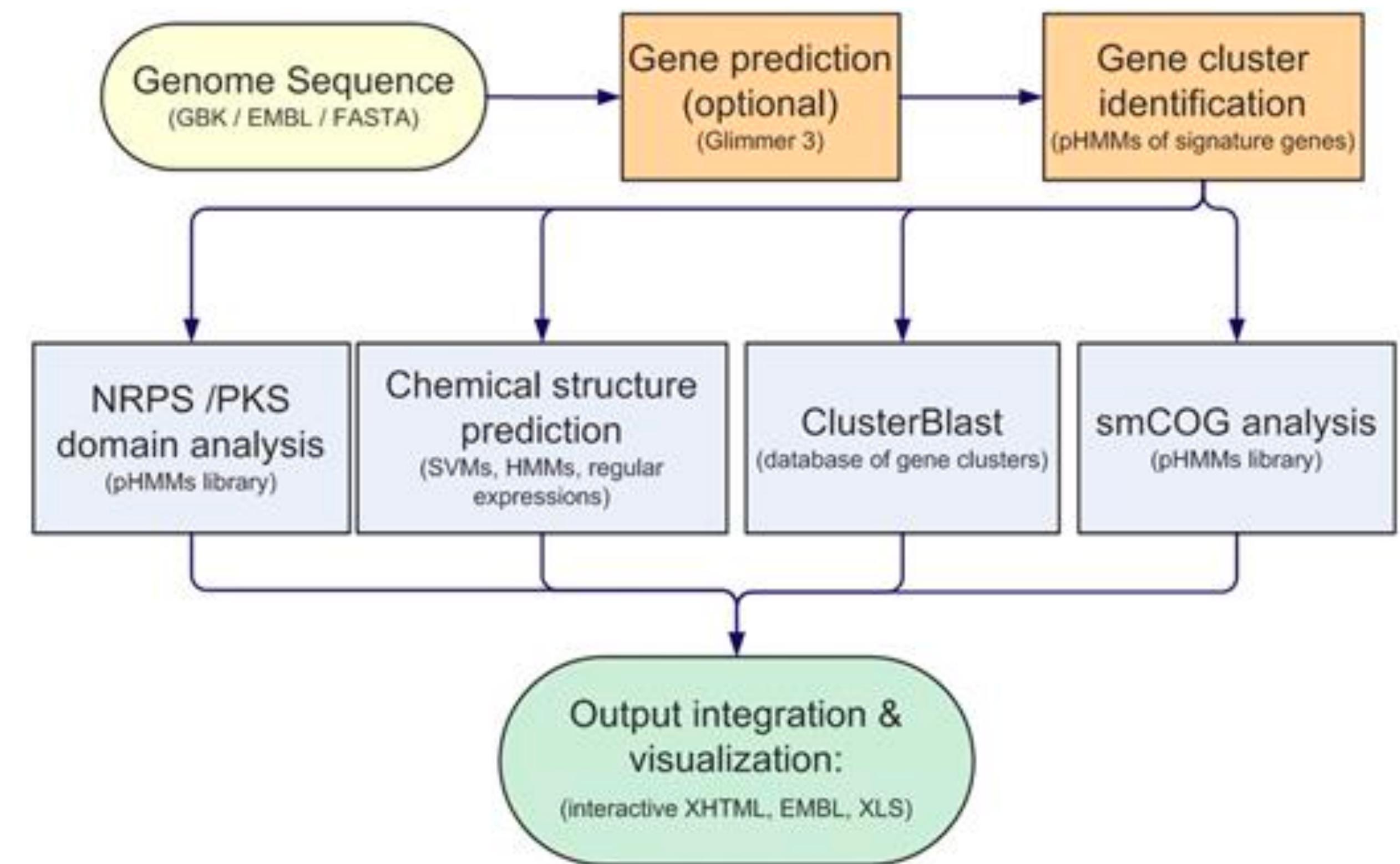


Please be considerate in your use of antiSMASH. Help us keep antiSMASH available for everybody by limiting yourself to 5 concurrent jobs. Need to run more? See the [antiSMASH install guide](#) for instructions for getting your own antiSMASH installation.



If you have found our SMASH useful, please [cite us](#).





Outline of the pipeline for genomic analysis of secondary metabolites. Genes are extracted or predicted from the input nucleotide sequence, and gene clusters are identified with signature gene pHMMs. Subsequently, several downstream analyses can be performed: NRPS/PKS domain analysis and annotation, prediction of the core chemical structure of PKSs and NRPSs, ClusterBlast gene cluster comparative analysis, and smCOG secondary metabolism protein family analysis. The output is visualized in an interactive XHTML web page, and all details are stored in an EMBL file for additional analysis and editing in a genome browser. A Microsoft Excel file with an overview of all detected gene clusters and their details is also generated.



anti
SMASH

fungi
SMASH

microbial genome
sequence



Detection

default

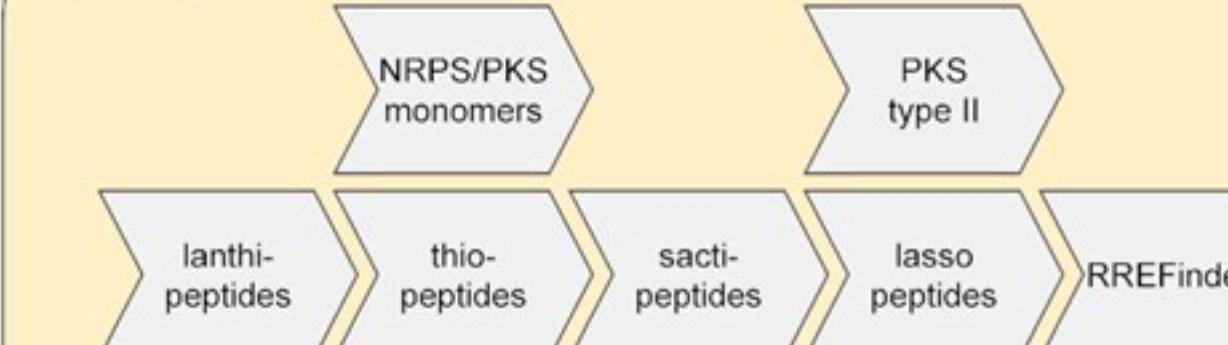


optional

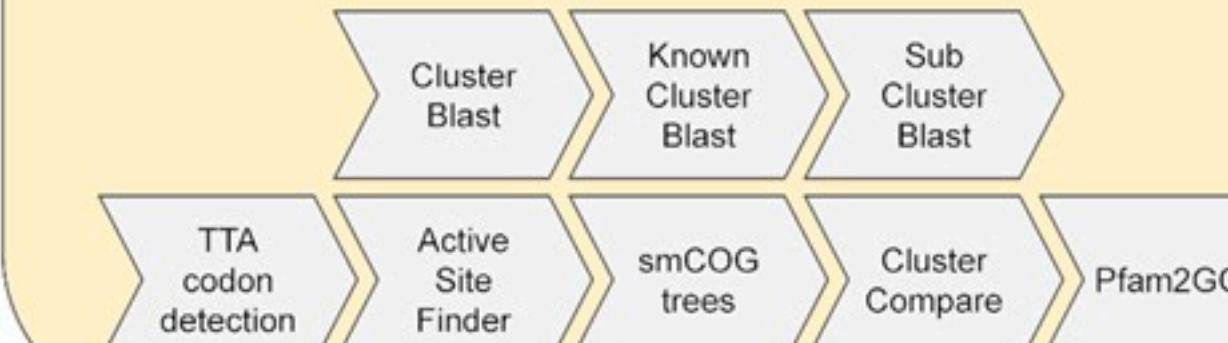


Analysis

type dependent



type independent



Output



<https://antismash.secondarymetabolites.org/#!/start>

antiSMASH 6.0: improving cluster detection and comparison capabilities

Kai Blin, Simon Shaw, Alexander M Kloosterman, Zach Charlop-Powers, Gilles P van Weezel, Marnix H Medema, & Tilmann Weber
Nucleic Acids Research (2021) doi: 10.1093/nar/gkab335.



← → ⌂ antismash.secondarymetabolites.org/upload/example/index.html

antiSMASH version 6.0.1

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

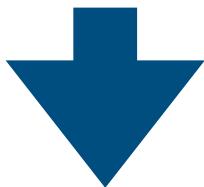
Identified secondary metabolite regions using strictness 'relaxed'

NC_003888.3 (Streptomyces coelicolor A3(2))

Region Type From To Most similar known cluster Similarity

Region	Type	From	To	Most similar known cluster	Similarity
Region 1	hglE-KS, T1PKS	86,637	139,654	leinamycin	NRP + Polyketide:Modular type I + Polyketide:Trans-AT type I 2%
Region 2	terpene	166,891	191,654	isorenieratene	Terpene 100%
Region 3	lantipeptide-class-i	246,868	270,397		
Region 4	NRPS	494,260	544,087	coelichelin	NRP 100%
Region 5	RiPP-like	791,701	799,942	informatipeptin	RiPP:Lanthipeptide 42%
Region 6	T3PKS	1,258,218	1,297,040	herboxidiene	Polyketide 8%
Region 7	ectoine	1,995,500	2,005,898	ectoine	Other 100%
Region 8	melanin	2,939,306	2,949,875	istamycin	Saccharide 4%
Region 9	siderophore	3,034,632	3,045,603	desferrioxamin B / desferrioxamine E	Other 83%
Region 10	NRPS	3,524,828	3,603,907	CDA1b / CDA2a / CDA2b / CDA3a / CDA3b / CDA4a / CDA4b	NRP:Ca+-dependent lipopeptide 87%
Region 11	T2PKS	5,496,474	5,567,376	actinorhodin	Polyketide:Type II 100%
Region 12	terpene	5,671,275	5,691,836	albaflavenone	Terpene 100%
Region 13	T2PKS	5,751,945	5,824,487	spore pigment	Polyketide 66%
Region 14	siderophore	6,336,091	6,346,368		
Region 15	NRPS-like, T1PKS, prodigiosin	6,430,010	6,475,291	undecylprodigiosin	NRP + Polyketide 100%

Genomes



Identification of secondary metabolites biosynthetic gene clusters



← → ⌂ antiSMASH version 6.0.1 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

Identified secondary metabolite regions using strictness 'relaxed'

NC_003888.3 (Streptomyces coelicolor A3(2))

Region	Type	From	To	Most similar known cluster	Similarity
Region 1	hglE-KS, T1PKS	86,637	139,654	leinamycin	NRP + Polyketide:Modular type I + Polyketide:Trans-AT type I 2%
Region 2	terpene	166,891	191,654	isorenieratene	Terpene 100%
Region 3	lantipeptide-class-i	246,868	270,397		
Region 4	NRPS	494,260	544,087	coelichelin	NRP 100%
Region 5	RiPP-like	791,701	799,942	informatipeptin	RiPP:Lanthipeptide 42%
Region 6	T3PKS	1,258,218	1,297,040	herboxidiene	Polyketide 8%
Region 7	ectoine	1,995,500	2,005,898	ectoine	Other 100%
Region 8	melanin	2,939,306	2,949,875	istamycin	Saccharide 4%
Region 9	siderophore	3,034,632	3,045,603	desferrioxamin B / desferrioxamine E	Other 83%
Region 10	NRPS	3,524,828	3,603,907	CDA1b / CDA2a / CDA2b / CDA3a / CDA3b / CDA4a / CDA4b	NRP:Ca+-dependent lipopeptide 87%
Region 11	T2PKS	5,496,474	5,567,376	actinorhodin	Polyketide:Type II 100%
Region 12	terpene	5,671,275	5,691,836	albaflavenone	Terpene 100%
Region 13	T2PKS	5,751,945	5,824,487	spore pigment	Polyketide 66%
Region 14	siderophore	6,336,091	6,346,368		
Region 15	NRPS-like, T1PKS, prodigiosin	6,430,010	6,475,291	undecylprodigiosin	NRP + Polyketide 100%

Genomes



Identification od secondary metabolites biosynthetic gene clusters



How to systematically compare those BGCs?



Biosynthetic Gene Similarity Clustering and Prospecting Engine

← → ⌂ bigscape-corason.secondarymetabolites.org ⌂ ⌂ ⌂ ⌂ ⌂

BIGSCAPE

HOME | INSTALLATION | TUTORIAL | ABOUT |

CORASON

BiG-SCAPE/CORASON pipeline
BiG-SCAPE and CORASON provide a set of tools to explore the diversity of biosynthetic gene clusters (BGCs) across large numbers of genomes, by constructing BGC sequence similarity networks, grouping BGCs into gene cluster families, and exploring gene cluster diversity linked to enzyme phylogenies.

Here you can find instructions for [installation](#) and a [BiG-SCAPE/CORASON tutorial](#). To get a more detailed explanation about parameters in BiG-SCAPE or CORASON please consult their wiki sites:

[BiG-SCAPE wiki](#) [CORASON wiki](#)

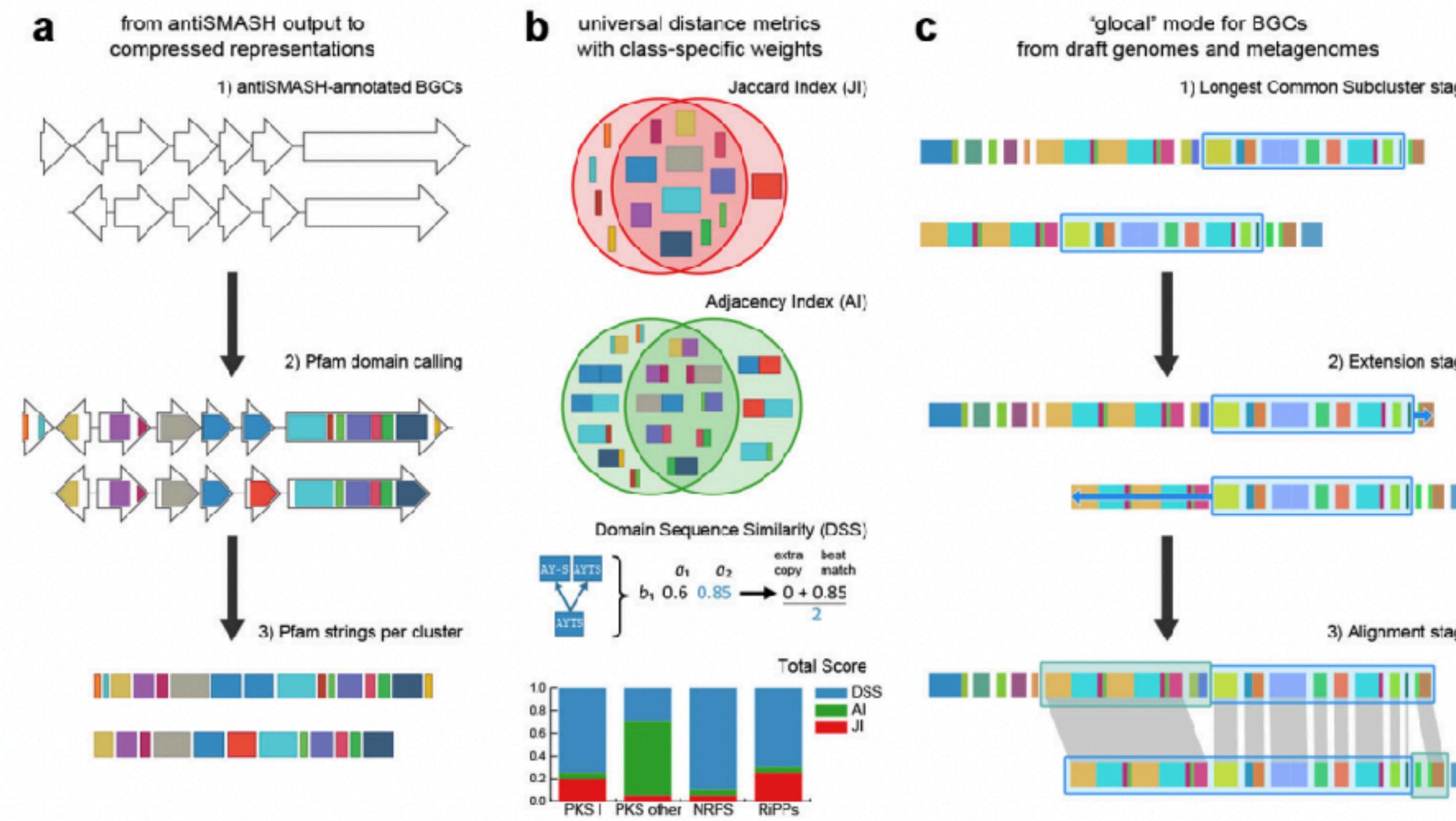


Fig. 2 **1. Main concepts in the BiG-SCAPE algorithm.**

a, Input data consists of BGC sequences directly imported from antiSMASH runs and/or from MIBiG. Nucleotide sequences are translated and represented as strings of Pfam domains. **b**, The three metrics that are combined in a single distance include the Jaccard Index (JI), which measures the percentage of shared types of domains; the Adjacency Index (AI), which measures the percentage of pairs of adjacent domains; and the Domain Sequence Similarity (DSS), which is a measure of sequence identity between protein domains encoded in BGC sequences. Weights of these indices have been optimized separately for different BGC classes. For simplicity, only four classes are shown. **c**, In “glocal” mode, BiG-SCAPE starts with the longest common subcluster of genes between a pair of BGCs and attempts to extend the selection of genes for comparison.