

Mining and Expressing Biosynthetic Gene Clusters from Soil Metagenomes

David Mead¹, Alinne Pereira², Megan Sandoval-Powers², Jinglie Zhou², Scott Monsma³, Kristin Degnes⁴, Alexander Wentzel⁴ and Mark R. Liles²

¹Varigen Biosciences; ²Auburn University; ³Lucigen Corp.; ⁴SINTEF

Abstract

Soil microorganisms contain vast reservoirs of bioactive natural products; however, the majority of them are recalcitrant to laboratory cultivation. In this study a large-insert soil metagenomic clone library (~110 kb and 19,200 clones) was constructed from an agricultural soil using a broad host range shuttle BAC vector. Clones containing biosynthetic gene clusters (BGCs) were identified using a 3D NGS pooling method in which plates, rows and columns were separately pooled and sequenced. Contigs were assembled from each pool and bioinformatically screened for secondary metabolite gene clusters using antiSMASH4.0. 474 clones containing a PKS and/or non-ribosomal peptide synthetase pathway among 1,516 total biosynthetic pathways were identified. These pathways are very divergent from known clusters, with the %G+C content varying from 34 to 79% and the nearest BLAST hit of keto-synthase domains ranging from 19 to 95% amino acid identity. New clades of keto-synthase domains were also found. 439 pathway-containing BAC clones were conjugally transferred into *Streptomyces coelicolor* M1154 and screened for the synthesis of antibacterial compounds against methicillin-resistant *Staphylococcus aureus*, *Acinetobacter baumannii*, the fungal pathogen *Candida albicans*, or human cancer cell lines, followed by LC/MS characterization of active clones to identify metabolites with bioactivity. The antibiosis hit rate was >10% for the bacterial and fungal screens. These results indicate that highly novel biosynthetic clusters can be cloned intact from complex metagenomes and heterologously expressed to produce secondary metabolites, thereby expanding our available resources or natural product discovery.

Three Libraries for Natural Product Pathways & Compounds

- Soil Diversity 1:** A BAC library with 19,200 clones from agricultural soil has been extensively characterized and validated to contain ~110,000 kb inserts with hundreds of full length PKS pathways. Every clone has been sequenced.
- Soil Diversity 2:** A BAC library with 103,680 clones from a native prairie soil has been partially characterized and validated to contain ~100,000 kb inserts.
- A PKS/NRPS enriched library** (439 clones) from the agricultural soil that has been shuttled to *Streptomyces* for antibiosis activity testing. Every clone has been sequenced.

Library	BAC Clones	Average Insert Size	BAC Vector	Insert Sequence	Anti-Microbial Activity	Known Structures
Soil Diversity 1	19,200	110 Kb	pSMART BACS	Yes	87 clones	26
Soil Diversity 2	103,680	100 Kb	pBAC-SBO	No	NA	NA
PKS/NRPS Express	439	100 Kb	pSMART BACS	Yes	59 Clones	NA

NGS results for Soil Diversity 1 metagenomic library reveals hundreds of BGCs encoding natural products

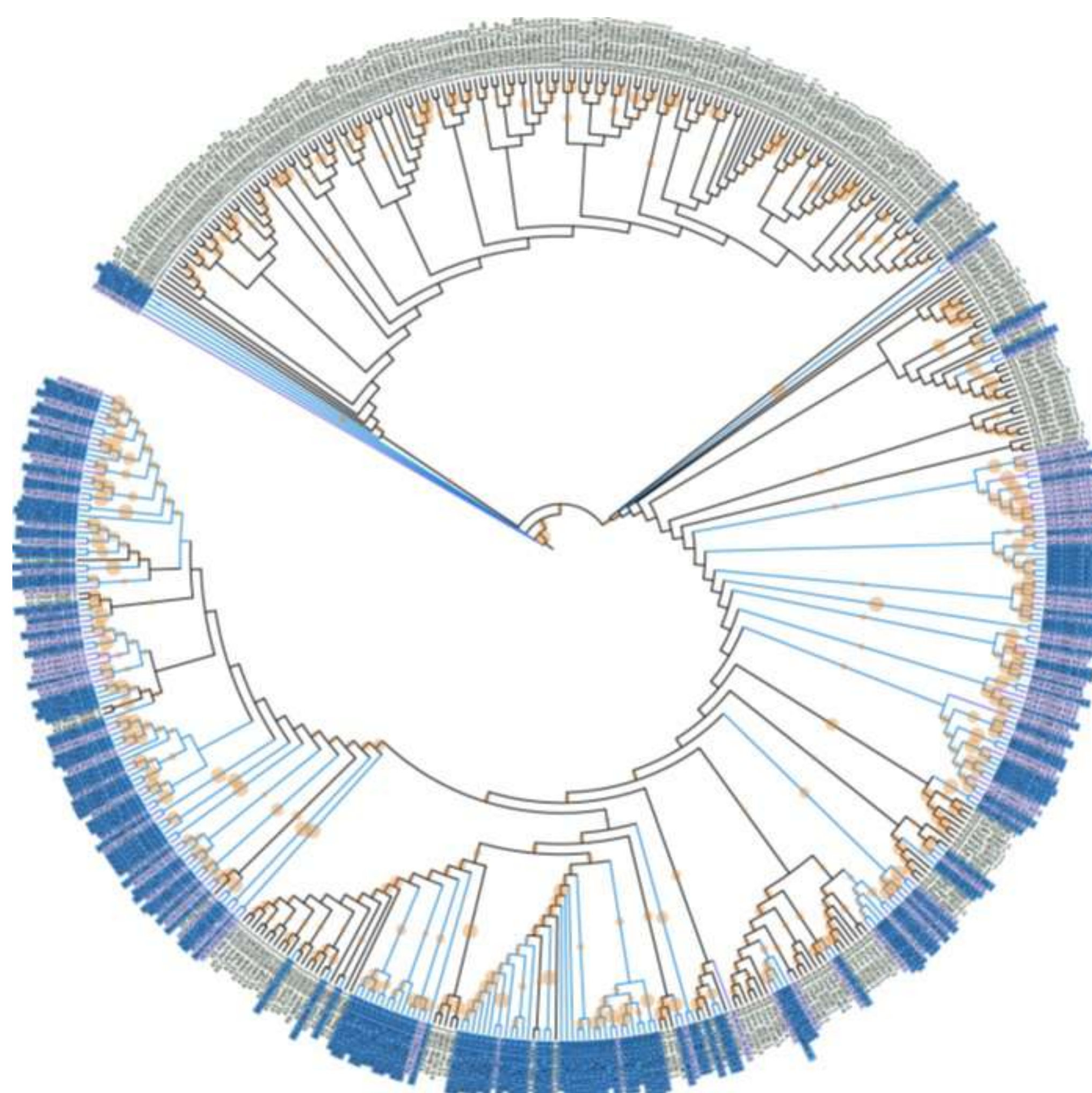
Cluster Type	PCR Screening	NGS Screening
NRPS	14	333
Terpene		306
Other	10	195
Bacteriocin		169
Type III PKS	5	97
Type I PKS	19	83
Arylpolyene		56
Type I PKS-NRPS	34	53
Lasso peptide	1	52
Lantipeptide	1	47
Resorcinol		22
Ladderane		21
Other KS	1	16
Hserlactone		15
Phosphonate		12
Type II PKS		8
Indole		7
Siderophore		5
Microviridin		4
Bacteriocin-NRPS		3
Butyrolactone		3
Cyanobactin		2
linaridin		2
Ectoine-terpene		1
Oligosaccharide		1
Phenazine		1
Phosphoglycolipid		1
Thiopeptide		1
Total	86	1516

16S rRNA Genes in 925 BACs

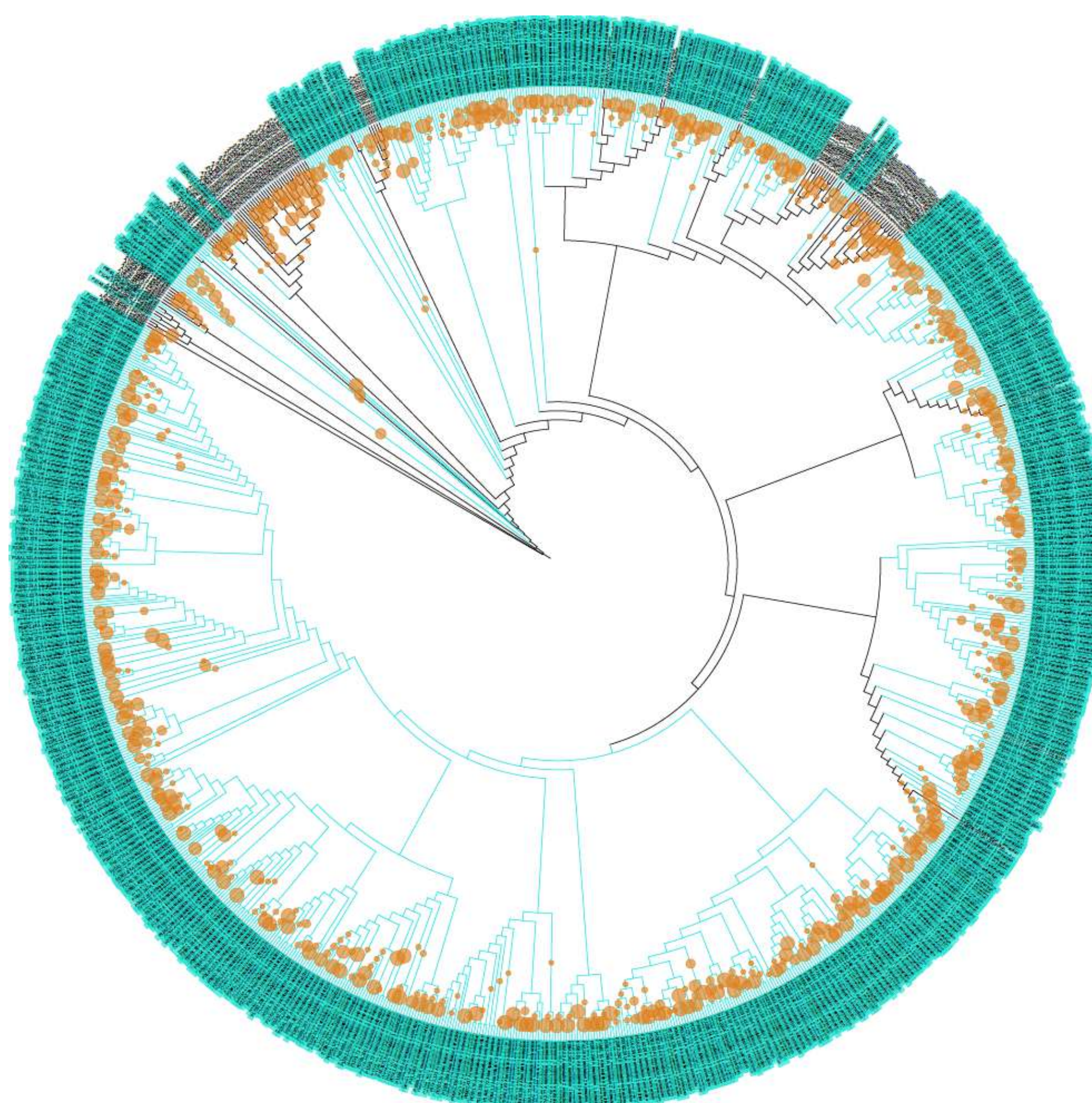
Clone	Contig length	Accession	Top hit	% Identity
21B7	106,238	CP007268	<i>Halorhodospira halochloris</i> str. A	84.46
21K7	155,294	FJ625356	Uncultured bacterium clone HF_NC_17	87.96
24N15	21,183	JQ371560	Uncultured bacterium clone FL3Aa9_7480	92.33
24O4	22,253	JX120409	Uncultured bacterium clone UA_39	95.12
31F6	12,075	CP007207	<i>Flavobacterium psychrophilum</i> FPG3	92.27
33E21	29,532	CP007035	<i>Niabella soli</i> DSM 19437	88.34
35L8	81,829	CP002826	<i>Oligotropha carboxidovorans</i> OM5	82.23
46F13	28,292	JX120409	Uncultured bacterium clone UA_39	95.12
46N9	30,069	AF507711	Uncultured soil bacterium clone S098	96.16
48M7	64,008	KC172332	Uncultured soil bacterium clone DM8-64	97.44
50F24	85,939	GQ396906	Uncultured bacterium clone AK4AB1_09G	97.06

16S rRNA containing contigs from 925 BACs were identified by BLAST search of the soil metagenomic library using *E. coli* 16S rRNA sequence as the query. 17 contigs were identified, corresponding to 11 unique clones. Each clone was then BLASTed against NCBI, and top hits were identified. Note that clone 35L8 (bold) also contains a Type I PKS gene cluster.

Novelty of KS and A domains from Metagenomic Pathways



Dendrogram of KS domains derived from all pathways encoding known polyketide products (black) along with metagenomic clones identified via PCR (purple) or NGS screening (blue). Domains in black are from known pathways with known structures.

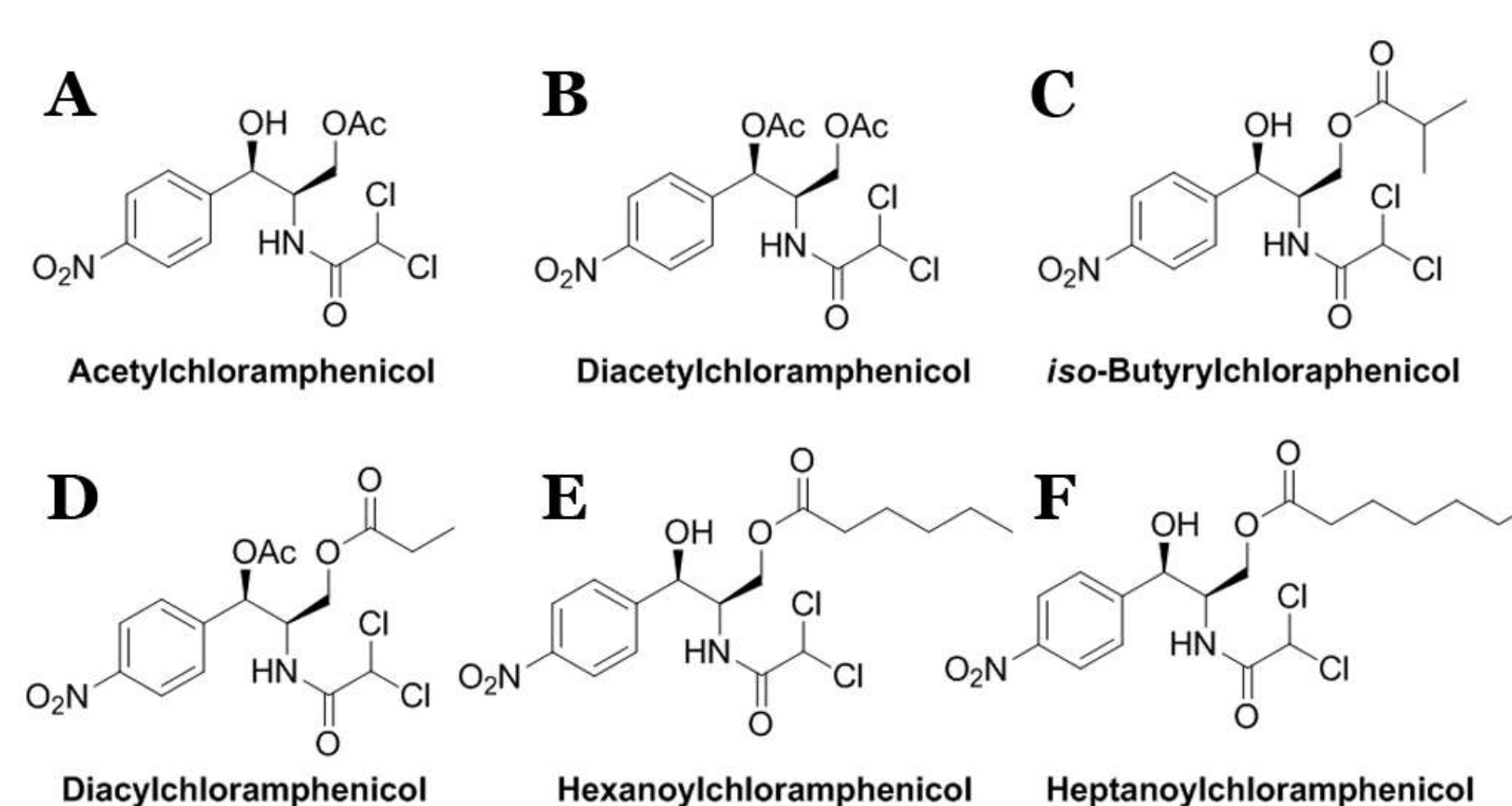


Dendrogram of A domains derived from NRPS pathways encoding known products (black) along with metagenomic clones identified via NGS (green). Domains in black are from known pathways with known structures.



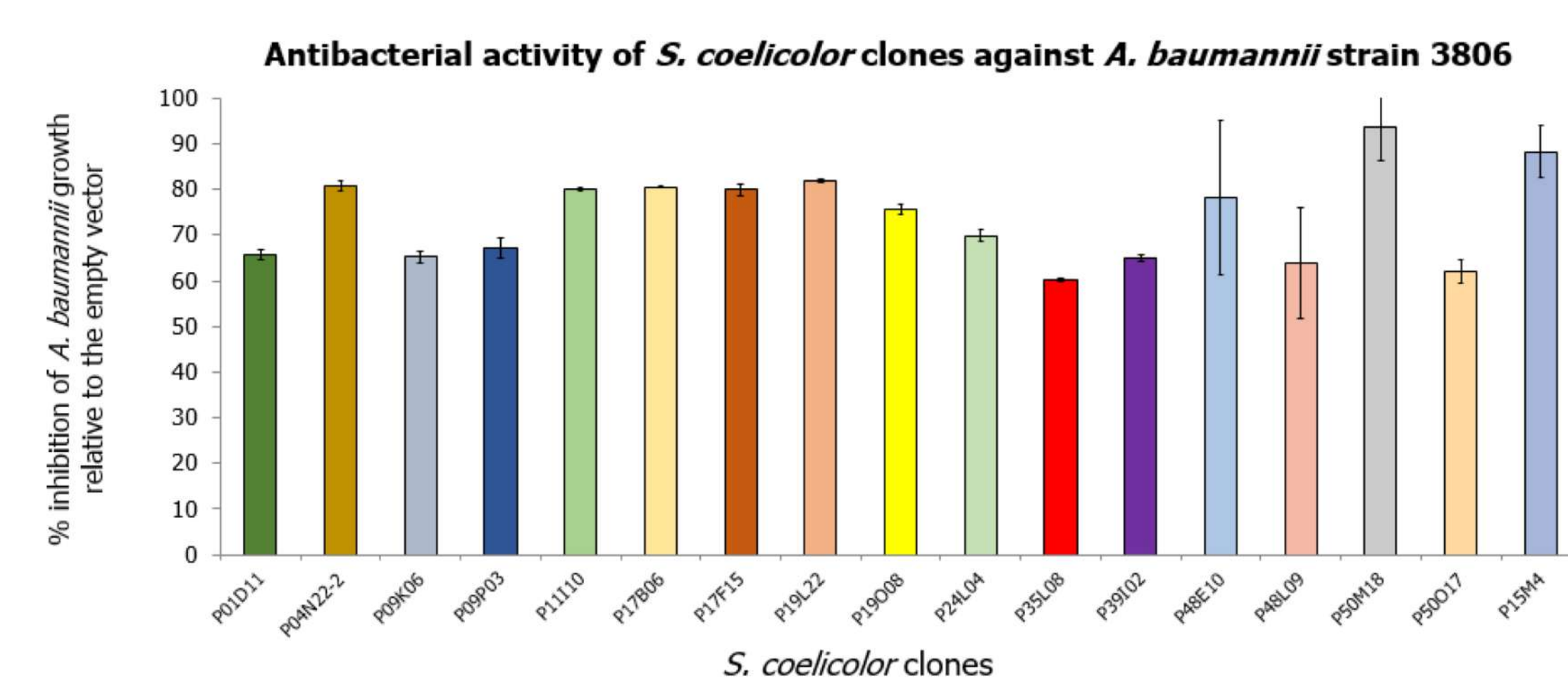
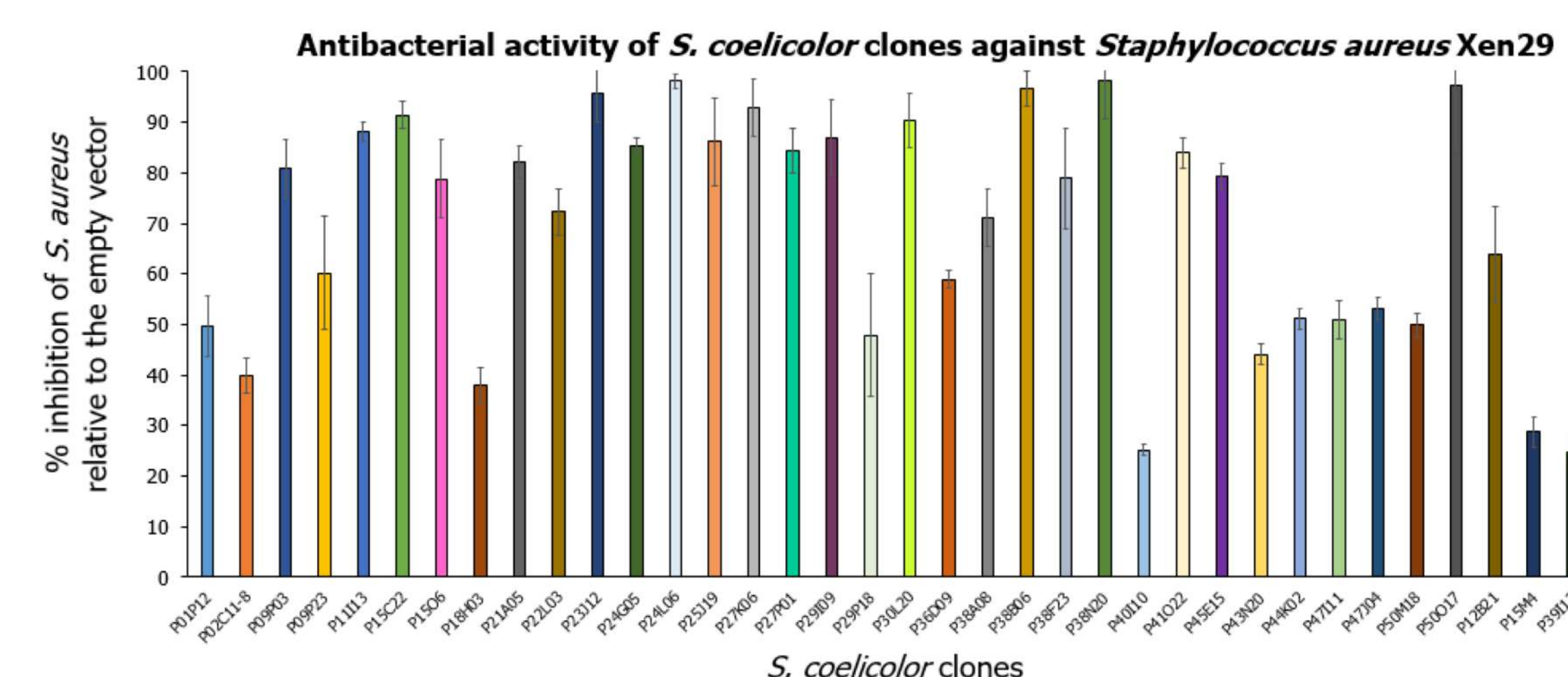
Chloramphenicol Derivatives with Antibacterial Activity Identified by Functional Metagenomics

Shamima Nasrin,^{1,2} Suresh Ganji,^{2,3} Kavita S. Kakirde,² Melissa R. Jacob,² Mei Wang,^{2,4} Ranga Rao Ravu,² Paul A. Cobine,² Ikhtlas A. Khan,^{2,3} Cheng-Cang Wu,² David A. Mead,¹ Xing-Cong Li,^{2,3,5} and Mark R. Liles^{2,3,5}

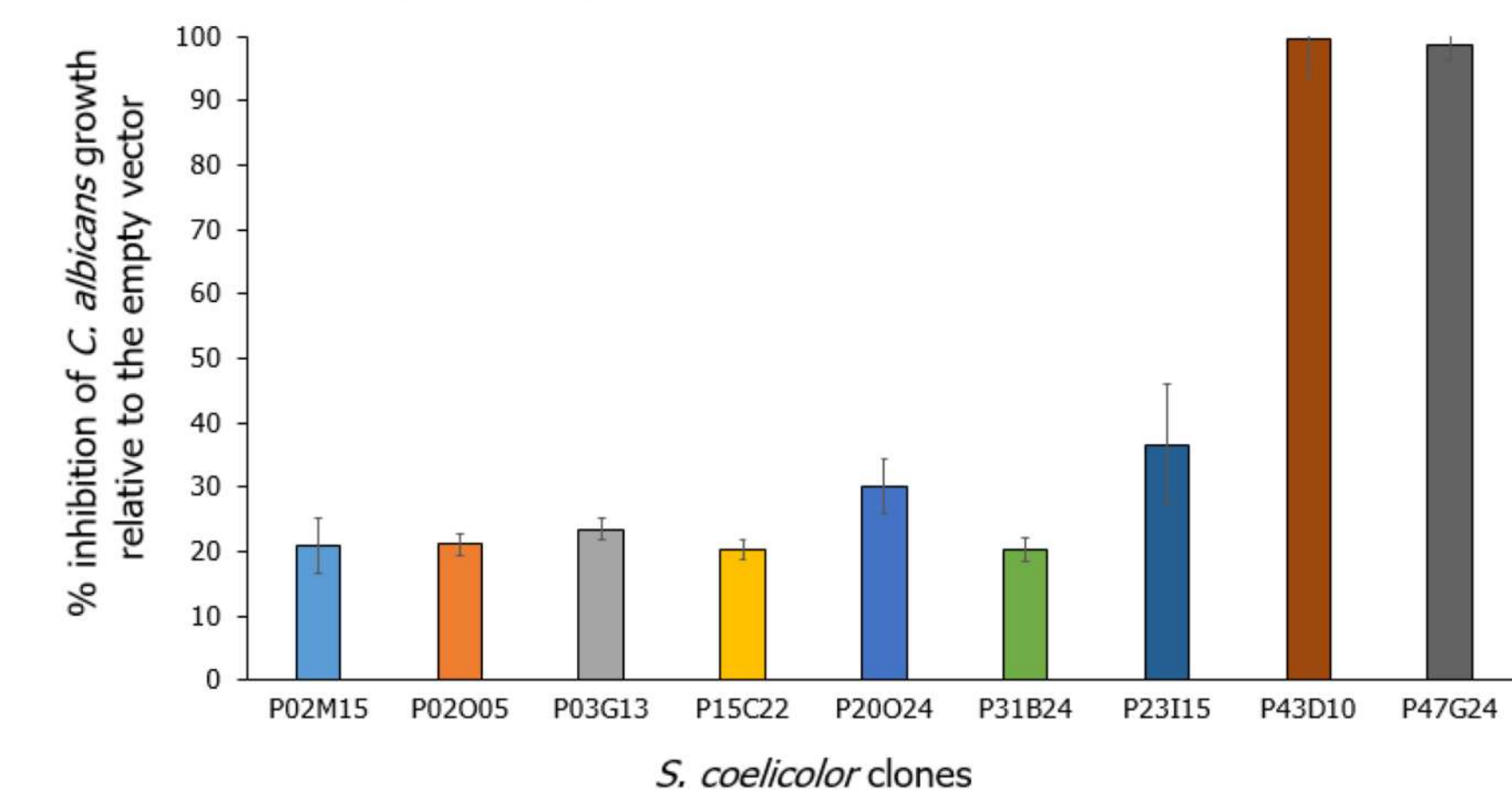


Structure of chloramphenicol derivatives identified in metagenomic clone supernatants based on LC/MS analyses. E and F are new compounds

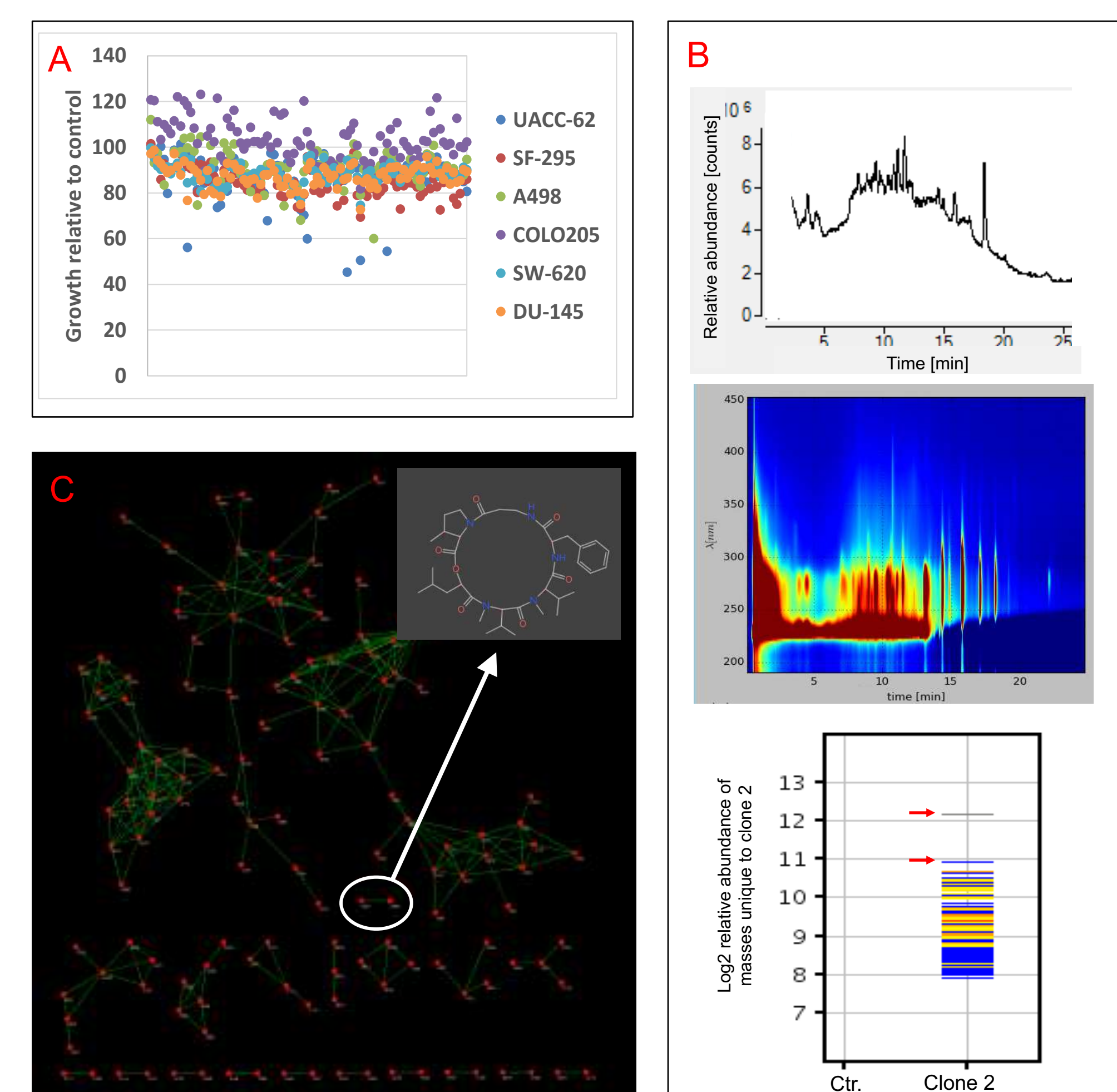
Antibiosis Activities from Metagenomic Clones



Antifungal activity of S. coelicolor clones against Candida albicans



EIGHT ANTI-CANCER CANDIDATE CLONES



A) Extracts of the PKS express library were tested for anticancer activity against six cell lines. B) Methanol extracts were analyzed with LC-DAD-MS, and molecules that were unique to clone 2 were identified by MS-profiling. C) The extract was re-analyzed with MSMS, and a molecular network with library search against compounds with similar MSMS pattern was generated. Further analysis remains, however one potential candidate mass was identified, and the MS pattern analysis indicated structural similarities to a cyclic peptide. The candidate from MS2 was in coherence with the MS1 profiling results.

ACKNOWLEDGEMENTS

This work was supported by NIH and NSF grants. Author Alinne Pereira is supported by the program Science without Borders – CAPES, Brazil. Work performed by SINTEF was supported by SINTEF and the Research Council of Norway, grant number 248885/O70