

Supplementary Material

Expanded natural product diversity revealed by analysis of lanthipeptide-like gene clusters in Actinobacteria

Qi Zhang, James R. Doroghazi, Xiling Zhao, Mark C. Walker, and Wilfred A. van der Donk*

Fig S1. Lanthipeptide gene cluster family similarity network for Actinobacteria.

Lanthipeptide gene cluster families discussed in this study are shown in blue, and gene cluster families with previously characterized members or close analogs are shown in red. GCFs are labeled with their respective numbers whereas gene clusters not belonging to families (i.e. singletons) are labeled with their cluster identifier. GCFs in grey are not discussed in this paper but can be found on the website (<http://www.igb.illinois.edu/labs/metcalf/gcf/lant.html>), together with their GCF numbers. To access the genes in the singleton clusters on the website, mouse over the singleton to show the GCF number. For instance, the very first singleton is Saci8_3. Then click Search at the top, and enter Saci8_3 in the search field for Gene Cluster Search. Then submit and click on the Saci8_3 link that appears.

Previously characterized lanthipeptides or their analogs are found in GCF31: SapB and various peptins (class III lanthipeptides); GCF22: actagardine (class II); GCF45: microbisporicin (class I); B24243_34 from *Streptomyces africanus* B-24243 is similar to cinnamycin (72% identical LanAs, class II); SACE_22: erythreapeptin from *Saccharopolyspora erythraea* NRRL 2338 (class III); FrEUN1fDRAFT_20 from *Frankia* sp. EUN1f is similar to cinnamycin (54% identical LanAs; class II).

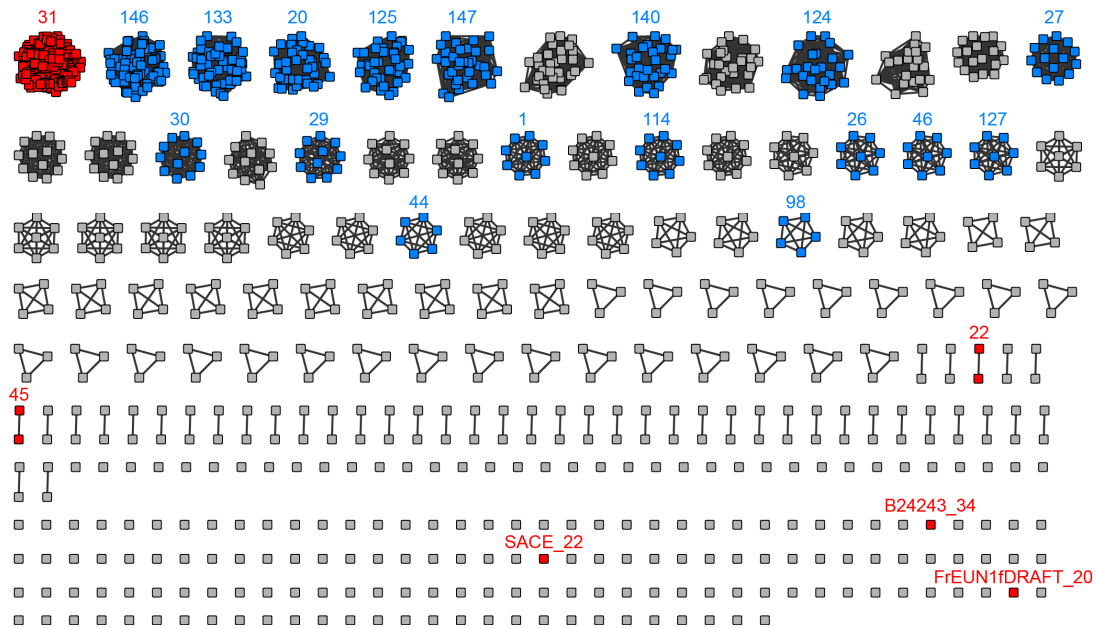


Fig S3. GCF46 that possibly encodes class III two-component lanthipeptides. (A) A representative gene cluster from GCF46. The two putative LanAs are labeled as 1 and 2. (B) Logo of LanA1 from GCF46. (C) Logo of LanA2 from GCF46.

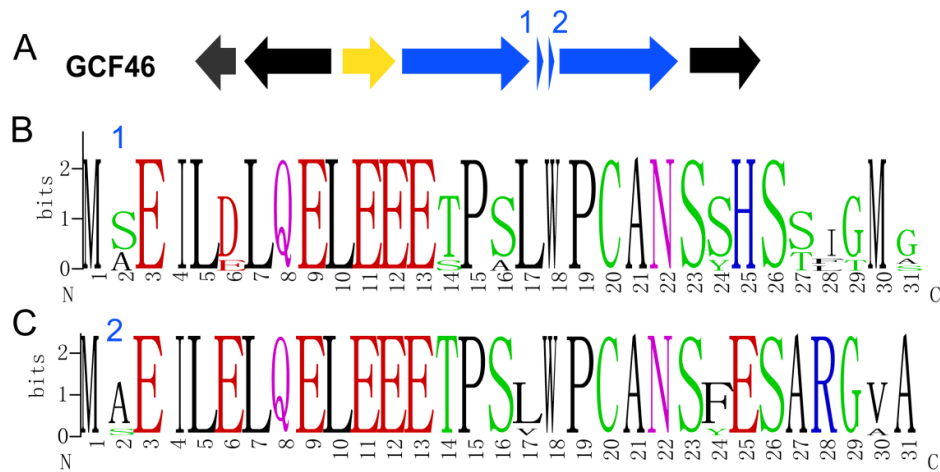


Fig S4. GCF140 that encodes a unique class of lanthipeptide-like biosynthetic systems.

(A) A representative gene cluster from GCF140. The three putative LanAs are labeled as 1, 2 and 3. (B) Logo of LanA1. (C) Logo of LanA2. (D) Logo of LanA3. The unusual Cys residue in the putative leader peptide region of LanA3 is highlighted by a red asterisk.

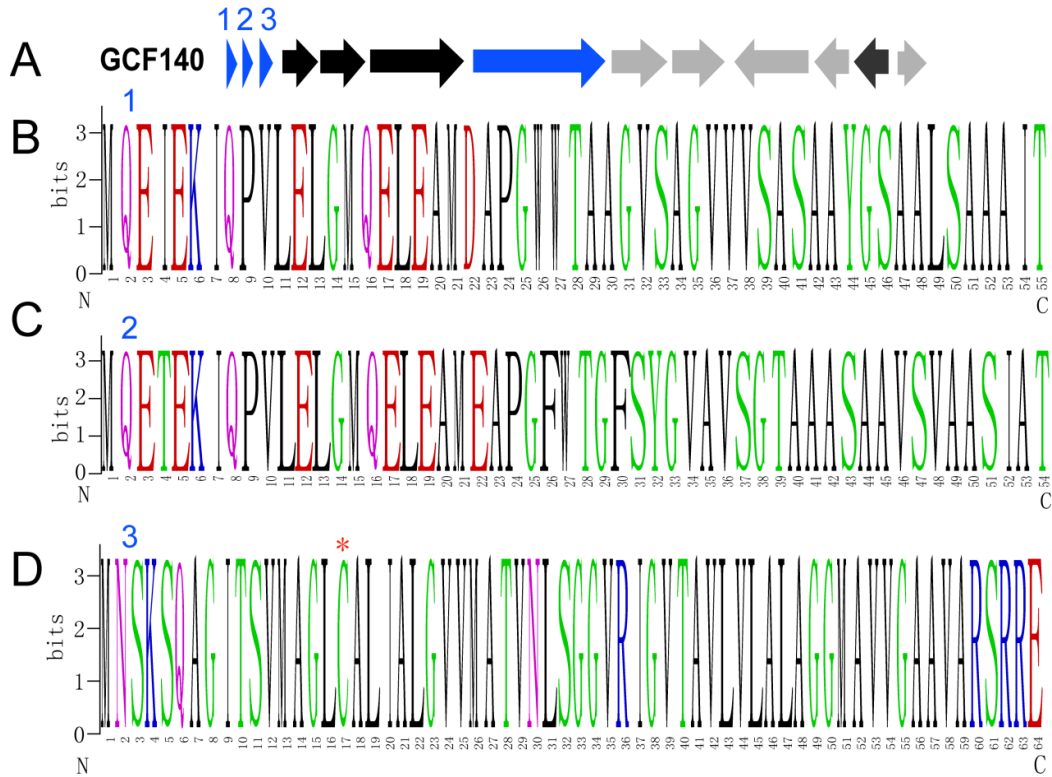


Fig S5. Sequence alignment of the putative LanAs in GCF26.

```

P357_6 : MDRHPNTTADA AV-----NAAEETDAAGGITLTKGRNRACARARVL AGMVLTSGLVITLTSLDTSVSAPS : 64
B2713_2 : MDQHLATTAADSVQ-----HSADEQDAAGGITLTKGRNRACARARVL AGMVLTSGLVIVTLSTLDTSVSAPG : 65
SHJG_12 : --MAELRRDASSAPADLPARPAADS-DAAGGITLTKGRNRACARARVL AGMVLTSGLVITLTSIDTSVSAPH : 68
B5429_14 : MDKNFAATAATADATEIAAEIAAEQDAAGGITLTKGRNRACARARVL AGMVLTSGLVITLSTIDTSVSAPR : 71
E11_28 : MDKNFAATAATADATEIAAEIAAEQDAAGGITLTKGRNRACARARVL AGMVLTSGLVITLSTMDTSVSAPR : 71
SSFG_18 : MDQHLTTTADTSVP-----SAAEQDAAGGITLTKGRNRACARARVL AGMVLTSGLVITLTSLDTSVSAPN : 65

```

P357 = *Streptomyces species* F-5065; B2713 = *Streptomyces hirsutus* B-2713; SHJG = *Streptomyces hygrosopicus* subsp. jinggangensis 5008; B5429 = *Streptomyces griseofuscus* B-5429; E11 = *Streptomyces species* F-6134; SSFG = *Streptomyces hirsutus* B-2713.

Fig S6. GCF26 zinc ligand residues and the residues that have replaced these ligands in a small subset of LanM proteins.

<i>Lactococcus lactis</i>	LctM	SQWCHGAS... ..FCLCHGILG...
<i>Prochlorococcus</i>	ProcM	ASWCHGAP... ..-HLCCGSLG...
<i>Streptomyces atratus</i>	Satratus_06001	PSWTTGLP... ..LSAGHGAFG...
<i>Streptomyces species F-5065</i>	P357_02363	STWSRGLA... ..LSLQGALG...
<i>Streptomyces hirsutus B-2713</i>	B2713_07235	VSWSGGFA... ..LSLGRGTLG...
<i>Streptomyces ghanaensis</i>		
ATCC 14672	SSFG_06280	ASWSRGLA... ..LSLQGTLG...
<i>Streptomyces species F-6134</i>	E11_05342	PSWAHGLA... ..LSLQGALG...
<i>Streptomyces griseofuscus B-5429</i>	B5429_04052	PSWAHGLA... ..LSLQGALG...
<i>Streptomyces species S-31</i>	P234_06451	PSWVSGTA... ..LSLQGTLG...
<i>Streptomyces hygrosopicus</i> subsp. jinggangensis 5008	SHJG_1658	LSWVSGLA... ..LSLQGALG...

Fig S8. MALDI-TOF-MS analysis of venezuelins produced by (A) *Streptomyces* species B-2375 and (B) *Streptomyces lavendulae* subspecies *lavendulae* B-2508.

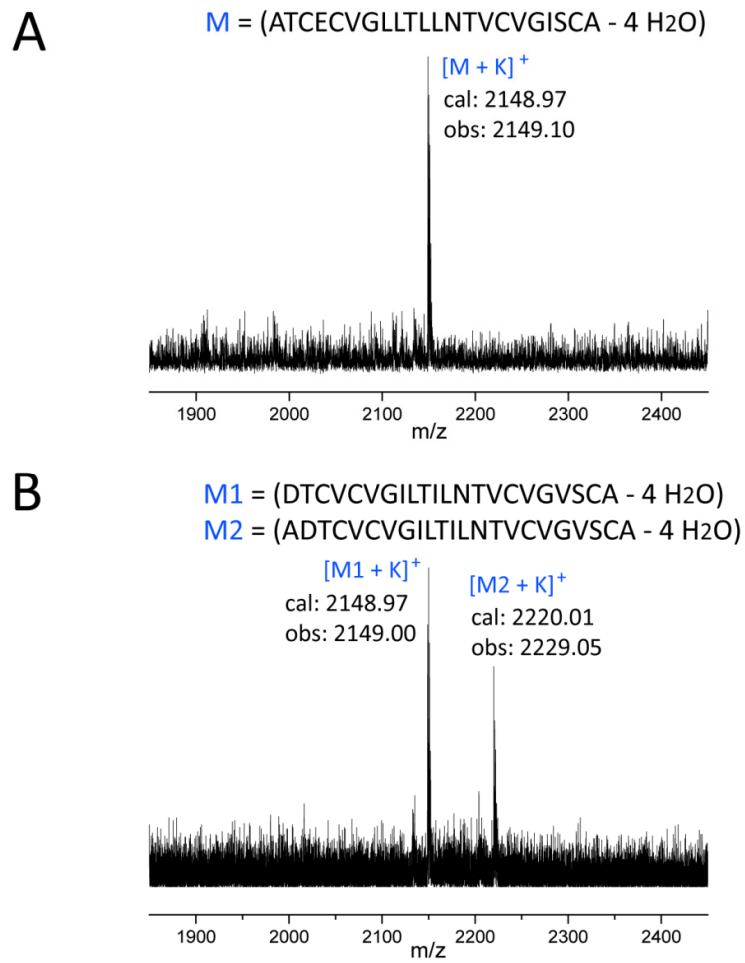


Table S1. Zinc ligands in LanM proteins in the genomes investigated in this study, or residues that have replaced these ligands in a small subset of LanM proteins.

Ligand residues	# of enzymes	Notes
CCH	196	Canonical Zinc ligand set of LanM proteins
CCC	3	Zinc ligand set found predominately in marine cyanobacteria
AGQ	2	Lant_GCF26
SGR	1	Lant_GCF26
SGQ	2	Lant_GCF26
VGQ	2	Lant_GCF26
TGH	1	Lant_GCF26
HHR	2	Lant_GCF98
HHH	1	Lant_GCF98
YRH	1	Lant_GCF98
AAH	1	1 LanM from Lant_GCF6, the others from GCF6 are CCH
ACA	2	TOMM_GCF77