SUPPORTING INFORMATION for

Automated genome mining of ribosomal peptide natural products

Hosein Mohimani[1], Roland D. Kersten[2], Wei-Ting Liu[3], Mingxun Wang[4], Samuel O. Purvine[5], Si Wu[5], Heather M. Brewer[5], Ljiljana Pasa-Tolic[5], Nuno Bandeira[4,6], Bradley S. Moore[2,6], Pavel A. Pevzner[4,*], Pieter C. Dorrestein[3,6,*]

[1] Department of Electrical and Computer Engineering, UC San Diego, La Jolla, California 92093, United States, [2] Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, UC San Diego, La Jolla, California 92093, United States, [3] Department of Chemistry and Biochemistry, UC San Diego, La Jolla, California 92093, United States, [4] Department of Computer Science and Engineering, UC San Diego, La Jolla, California 92093, United States, [5] Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington 99354, United States, [6] Skaggs School of Pharmacy and Pharmaceutical Sciences, UC San Diego, La Jolla, California 92093, United States. * contact pdorrestein@ucsd.edu or ppevzner@ucsd.edu

Materials and methods

Extraction of microbial metabolites. We obtained 16 *Streptomyces* strains described below in the Genome Datasets section. Strains were grown on ISP2 agar plates (4 g yeast extracts, 10 g malt extract, 4 g D-glucose, 18 g agar, ad 1000 ml water). Each agar plate was inoculated with each bacterial strain by 4 parallel streaks. The plates were incubated for 10 d at 28 °C. The agar was sliced into small pieces, covered with equal amount of Milli-Q water and n-butanol in a 50 ml centrifuge tube and shaken at 225 rpm for 12 h at 28 °C. The n-butanol layer was subsequently collected using transfer pipette and dried *in vacuo*.

Genome datasets. Genomes of 18 strains of *Streptomyces* were recently sequenced at Broad Institute and are available from the *Actinomycetales* database website.[1] Genomes of *Streptomyces griseus* IFO 13350 (AP009493) and *Streptomyces coelicolor*A3(2) (AL645882) are available from NCBI (**Table S1**).

Spectral datasets (CID) of microbial extracts. Collision-induced dissociation (CID) MS/MS datasets were collected with or without liquid chromatography (LC) separation in-line with mass spectrometry. For LC-MS, capillary columns were prepared by drawing a 360 µm O.D., 100 µm I.D. deactivated, fused silica tubing (Agilent) with a Model P-2000 laser puller (Sutter Instruments) (Heat: 330, 325, 320; Vel, 45; Del, 125) and were packed at 600 psi to a length of about 10 cm with C18 reverse-phase resin suspended in methanol. The column was equilibrated with 95% of solvent A (water, 0.1% AcOH) and loaded with 10 µl (10 ng/µl in 10% CH_3CN) of bacterial butanol extract by flowing 95% of solvent A and 5% of solvent B (CH_3CN, 0.1% AcOH) at 200 µl/min for 15 mins. A gradient was established with a time-varying solvent mixture [(min, % of solvent A): (20, 95), (30, 60), (75, 5)] and directly electrosprayed into the LTQ-FT MS inlet (source voltage, 1.8 kV; capillary temperature, 180 °C). The first scan was a high resolution broadband scan. The subsequent six scans were low resolution data-dependent on the first scan. In each data-dependent scan, the top intensity ions excluded the ones in exclusion list were selected to be fragmented by CID which generated hundreds of fragmentation spectra collected as individual data events. The resulting .RAW files were converted to .mzXML using the program ReAdW (http://tools.proteomecenter.org).

Spectral datasets (HCD) of microbial extracts. Higher-energy collisional dissociation (HCD) datasets were acquired from samples prepared in 20% acetonitrile before injection. The constant flow capillary RPLC system used for peptide separations was similar to the previous report.[2] Briefly, the HPLC system consisted of a custom configuration of Agilent 1200 nanoflow pumps (Agilent Technologies), 2-position Valco valves (Valco Instruments Co., Houston, TX), and a PAL autosampler (Leap Technologies, Carrboro, NC), allowing for fully automated sample

analysis across four separate HPLC columns (3-µm Jupiter C18 stationary phase, Phenomenex, Torrence, CA). Mobile phases consisted of 0.1% formic acid in water (A) and 0.1% formic acid acetonitrile (B). Flow rate through the capillary HPLC column was set as 300 nL/min. The HPLC system was equilibrated with 100% mobile phase A, and the following gradient was started 40 min after injection (5 µL sample loop): 0-2 min, 0-8% buffer B; 2-20 min, 8-12% buffer B; 20-75 min, 12-80% buffer B; 75-97 min, 80-95% buffer B. ESI using an etched fused-silica tip (*42*) was employed to interface the RPLC separation to a LTQ Orbitrap Velos mass spectrometer (Thermo Scientific, San Jose, CA). Precursor ion mass spectra (automatic gain control was set to $1x10^6$) were collected for 400-2000 m/z range at a resolution of 60K followed by data-dependent HCD MS/MS (resolution 7.5K, normalized collision energy 45%, isolation window 2.5 Th, activation time 0.1 ms, AGC $5x10^4$) of the ten most abundant ions. A dynamic exclusion time of 30 s was used to discriminate against previously analyzed ions.

Genome mining of lanthipeptide gene clusters. Lanthipeptides are encoded by a lanthipeptide structural gene (LanA). LanA is not a suitable candidate for genome mining since it is extremely variable across different bacteria. Since no algorithms for predicting LanA exist, we focus on other more conserved biosynthetic enzymes in the lanthipeptide gene cluster. RiPPquest searches for the more conservative LANC-like domain (Pfam: PF05147) and capitalizes on the observation that there exists a LANC-like domain in close vicinity of each LanA gene.

We analyzed a 10 kb window centered at the structural LanA gene of 22 known lanthipeptides (**Fig. S 2)** and considered the most conserved genes and corresponding Pfam domains in these windows. Some of the identified Pfam domains are conserved in specific classes, e.g. the Lant_dehyd_N and Lant_dehyd_C domains occurring only in class I lanthipeptides, the Peptidase_C39 domain occurring only in class II lanthipeptides, and the Pkinase domain occurring only in class III lanthipeptides. However, the LANC-like domain (LanM/LanC1), the ABC_membrane domain, and the ABC_tran domain occur in the selected 10 kb window of all different classes of lanthipeptides. We have selected the LANC-like domain for Pfam domain search in RiPPquest due to its higher specificity to lanthipeptides (**Fig. S2**).

For each LANC-like domain in the microbial genome, a window of 10 kb centered at this domain is selected to form a database of putative core lanthipeptides for follow up MS/MS database search. Since lanthipetides usually appear in short ORFs, we further restrict our analysis to ORFs < 100 aa in the 6-frame translation of the genome. Because core lanthipeptides always appear at the C-terminus of an ORF, we only consider the peptide sequence of the C-terminal half of an ORF. This reduction the database size in RiPPquest searches is important since lanthipeptides are often poorly fragmented and identification of such poorly fragmented spectra in searches against large databases is problematic.

As an example, the *Streptomyces roseosporus* NRRL 11379 genome (approximately 9 Mb) has three lanthipeptide gene clusters, with a total of 132 short ORFs <100 aa, including three ORFs producing lanthipeptides SRO-2212, SRO3108[3], and another hypothetical lanthipeptide producing ORF (**Fig. S3**). Because the LANC-like domain has between one to six hits to the genome of a *Streptomyces*, the database of putative core lanthipeptides is about 100 times smaller than for the whole *Streptomyces* genome.

Mass spectrometry analysis of lanthipeptide modifications. The most essential lanthipeptide modifications are dehydration of serine and threonine, and formation of the lanthionine and methyllanthionine bridges. Furthermore, a thiol elimination mechanism for lanthionine PTMs during mass spectrometry yields Cys and Dha at the position of Ser and Cys, respectively, in the core peptide[3]. **Fig. 1f,g** shows all possible modified (mature) peptides for a hypothetical core lanthipeptide Thr-Phe-Cys-Arg-Ser. From a mass spectrometry standpoint, there are eight possible products by accumulation of PTMs, resulting in six possible scenarios for observed mass shifts in mass spectrometry (allowing *Ser → Dha*, *Ser → Cys*, *Cys → Dha* and *Thr → Dhb*). The number of possible mature peptides increases exponentially with the number of serines, threonines and cysteines in the core peptide, making it time consuming to try all possible combinations of PTMs for every spectrum. For example, for the 22 aa core peptide of lanthipeptide SRO-2212 (TGSQVSLLVCEYSSLSVVLCTP), a total of 1088 possible mature peptides exist.

While RiPPquest is currently limited to lanthipetide analysis, it can be extended to the majority of other RiPP classes as soon as (i) it implements a biosynthetic rationale for transforming core into mature peptide for a specific RiPP class, and (ii) it implements a genome mining rationale for a specific RiPPs class.

Scoring peptide spectrum matches. All MS/MS database search tools score Peptide-Spectrum Matches (PSMs) with the goal to find out how well the experimental spectrum is explained by the theoretical spectrum formed by the fragment ions of the peptide (**Fig. 1i**). We have chosen to score PSMs using an advanced scoring function used in *de novo* peptide sequencing PepNovo[4]. In the brute force approach, one forms PSMs between each spectrum in the spectral dataset and each modified core peptide if the parent masses of the spectrum and the modified core peptide are close to each other (within 0.5 Da). In the case of lanthipeptide SRO-2212, 335 out of 1088 possible modifications of the core peptide are within 0.5 Da of the precursor mass of SRO-2212 (doubly charged 1107.04 m/z). Because it is time consuming to compare each spectrum against each possible modified peptide for large spectral datasets, we use the spectral alignment technique to efficiently find modifications of the core peptide that best matches the spectrum[5-8].

Converting scores to p-values. While PSM scores are useful for selecting top-scoring PSMs, they are notoriously unreliable for estimating the statistical significance of PSMs[9]. To convert scores into p-values, RiPPquest uses a recently developed MS-DPR approach for evaluating p-values of PSMs[10]. While other methods for evaluating p-values exist[11], MS-DPR is the only approach available today for evaluating p-values of PSMs formed by non-linear, e.g. cyclic peptides. Since many RiPPs are non-linear, estimating p-values via MS-DPR will be the only option when RiPPquest is extended from lanthipeptides to other non-linear RiPPs such as cyanobactins or lassopeptides.

Spectral networks. Spectral networks are a visualizion of spectra as familial groupings of corresponding peptides. Edges in a spectral network connect nodes corresponding to spectra that represent peptides differing from each other by a mutation or a modification. Such pairs of spectra connected by edges in the spectral network are revealed using the spectral alignment approach. Spectral networks enable discovery of novel homologs of known peptides, and novel families of related peptides. Most classes of RiPPs form families of related peptides, making spectral networks helpful in RiPP analysis. In particular, spectral networks reveal related lanthipeptides with stepwise *N*-terminal leader processing and different dehydration numbers.

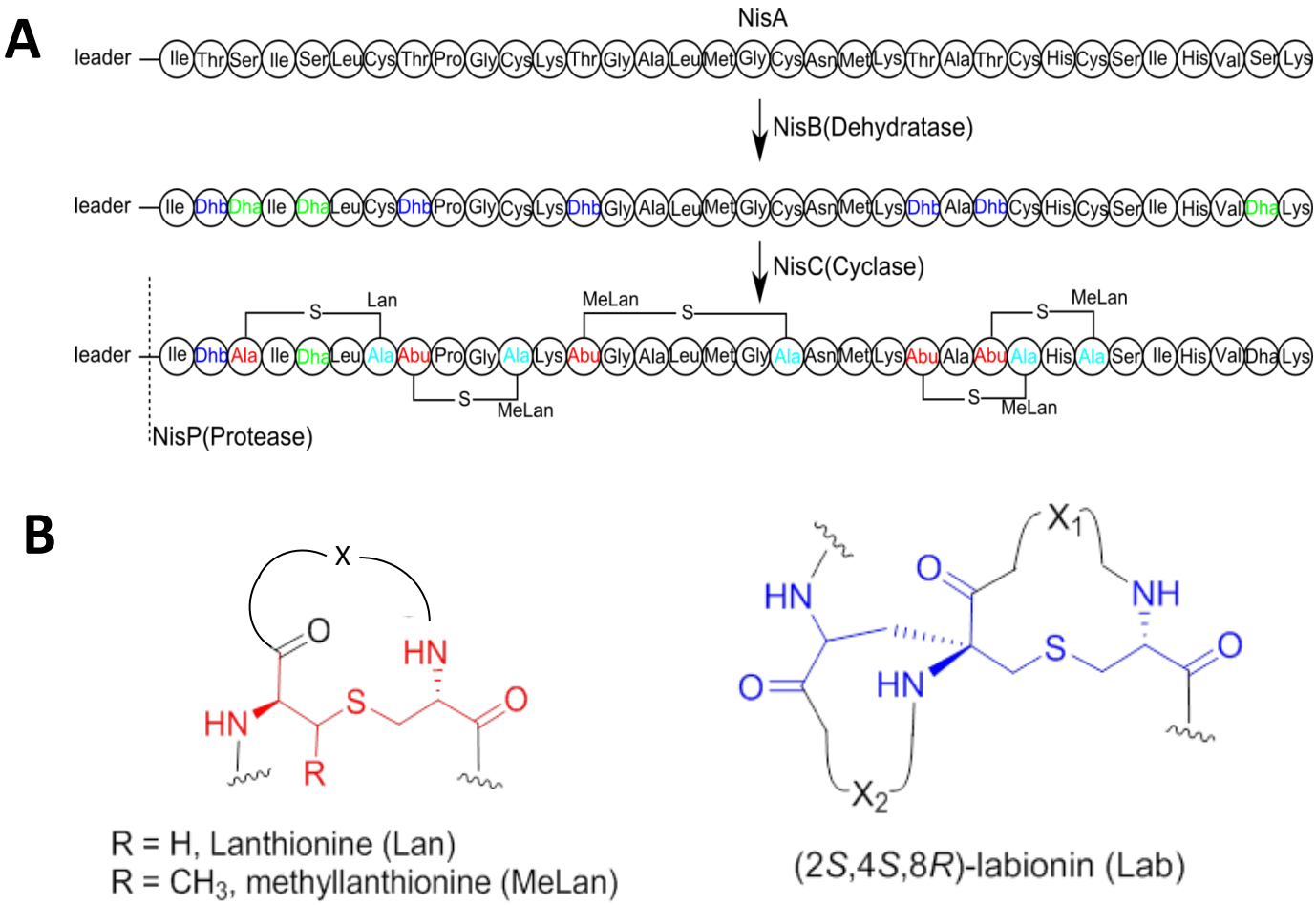


Figure S1. (A) Lanthipeptide biosynthesis, exemplified by nisin. The 34 aa core peptide of nisin is encoded as the C-terminus of a 57 aa precursor protein. A lanthionine dehydratase (NisB) transforms Ser and Thr residues of the core peptide NisA into Dha and Dhb residues, respectively. Cyclase NisC introduces Lan and MeLan residues by bridging Dha and Dhb residues, respectively, to a Cys residue. Finally, protease NisP cuts the modified precursor peptide and releases the lanthipeptide nisin from the leader peptide. (B) Structures of lanthionine, methyllanthionine and labionin.

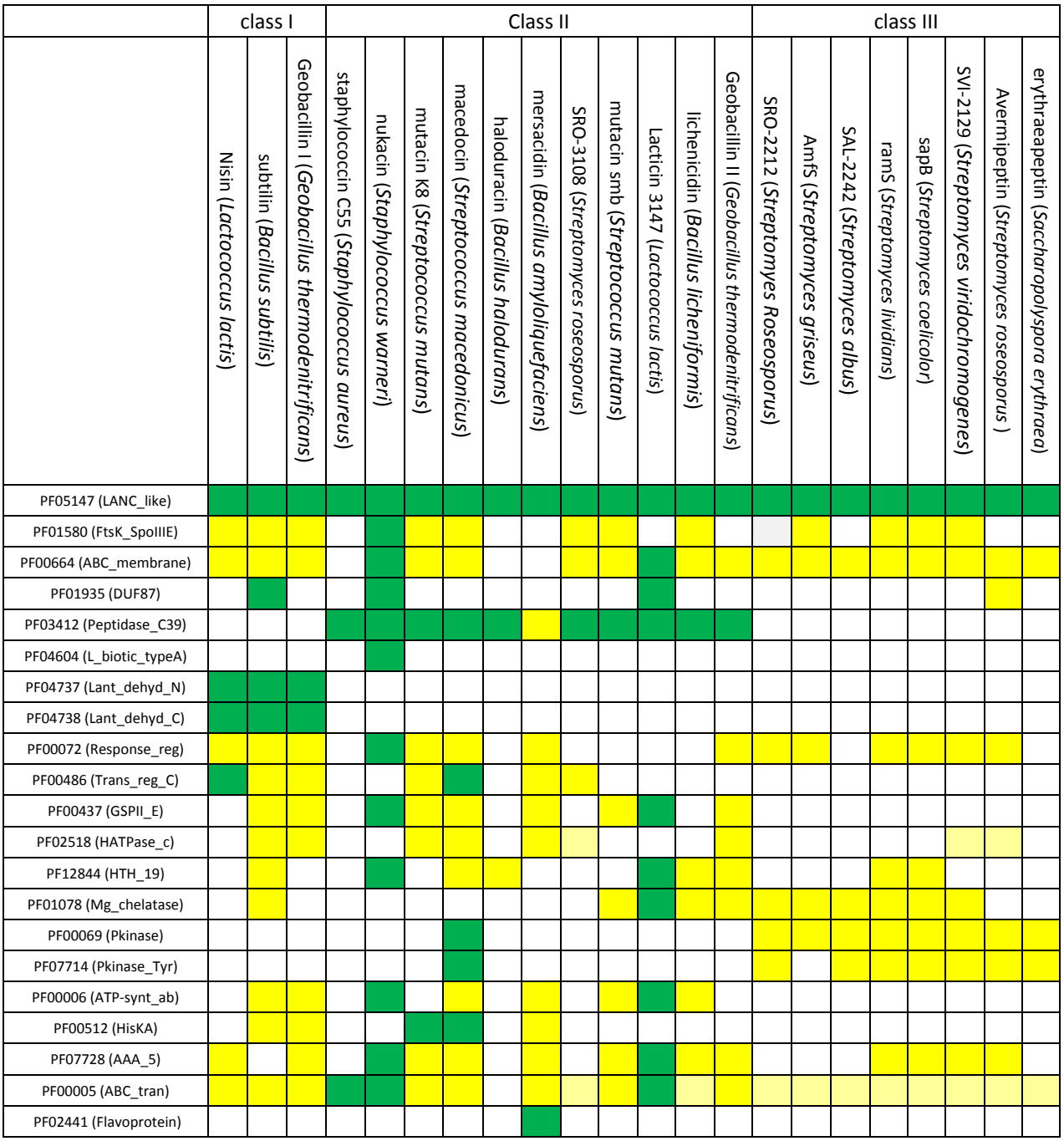
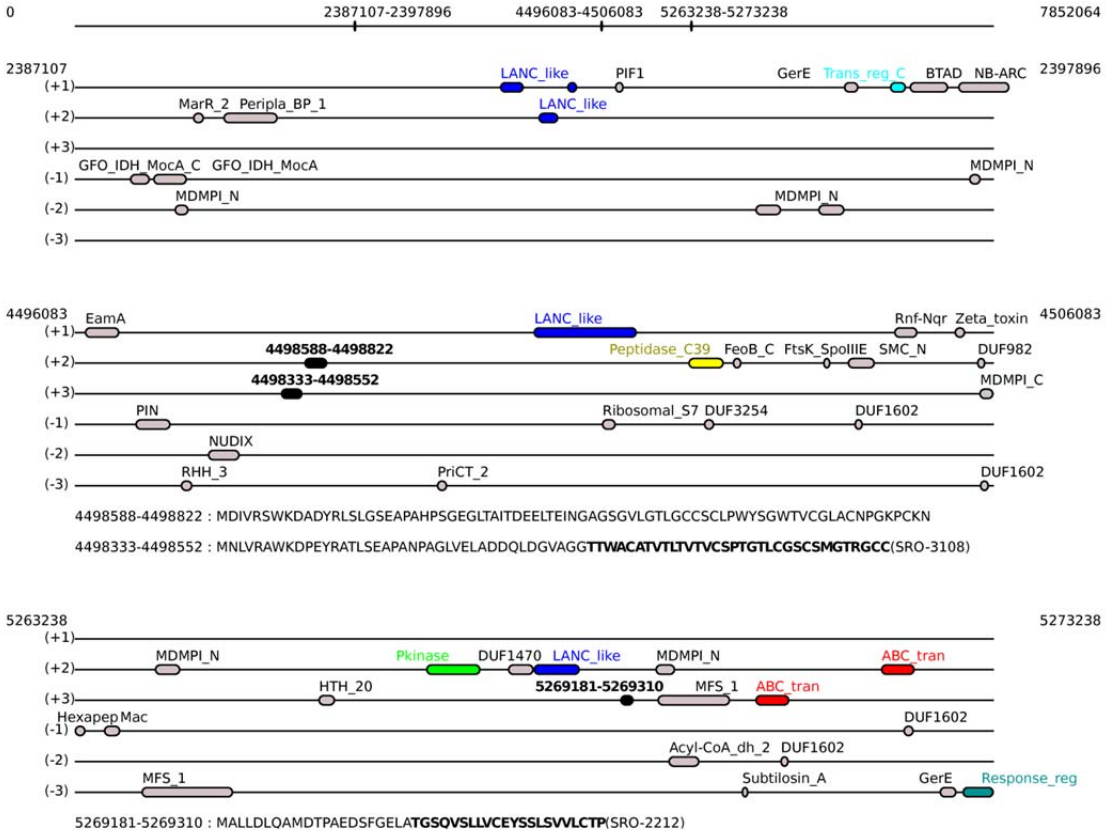


Figure S2. List of common Pfam domains in lanthipeptide gene clusters. A colored square means the corresponding Pfam domain occurs within a 10 kb window centered at precursor lanthipeptides. The color is green if the number of occurrences of the domain in the whole microbial genome is less than 10, dark yellow if the number of occurences is between 10-100, and light yellow if more than 100. Only lanthipeptide produced by microorganisms with known genome sequence are considered. While some of the Pfam domains are class specific (e.g. Lant_dehyd_N and Lant_dehyd_C occurring only in class I lanthipeptides, Peptidase_C39 occuring only in class II lanthipeptides, and Pkinase occurring only in class III lanthipeptides), LANC_like, ABC_membrane domain, and ABC_tran domain occur in all lanthipeptides, irrespective of their class. Among them LANC_like is preferred for lanthipeptide genome mining, due to its higher specificity to lanthipeptides.

A - *Streptomyces roseosporus* NRRL 11379



B - *Streptomyces roseosporus* NRRL 15998

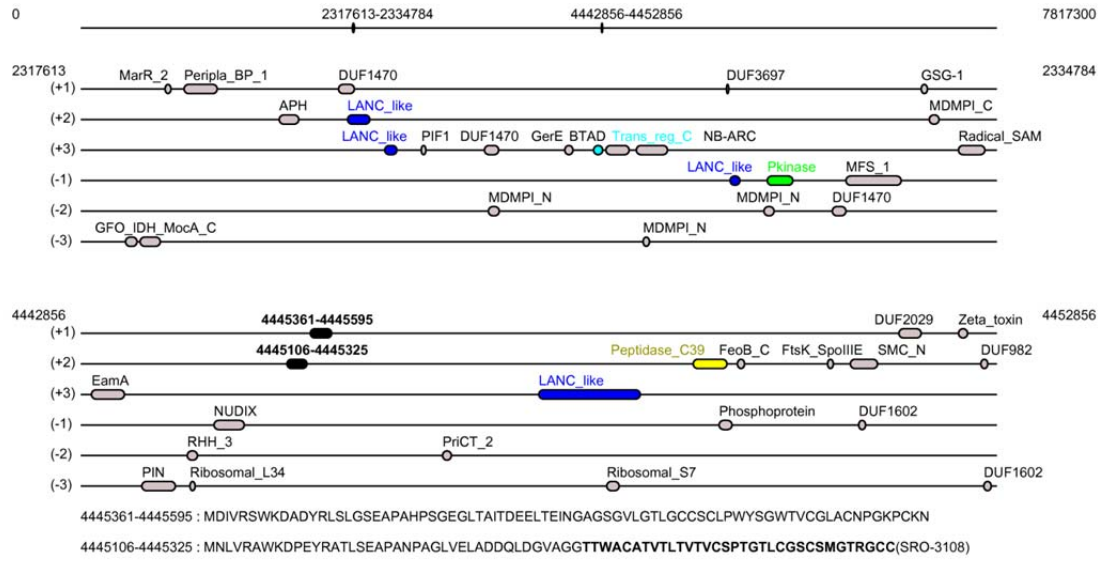


Figure S3. Lanthipeptide gene clusters predicted by RiPPquest in the genome of (A) *Streptomyces roseosporus* NRRL 11379, (B) *Streptomyces roseosporus* NRRL 15998. The figure shows all Pfam domains discovered in a 10,000 bp window centered at LANC_like domain, and predicted lanthipeptide precursor ORFs.

C – *Streptomyces* sp. AA4

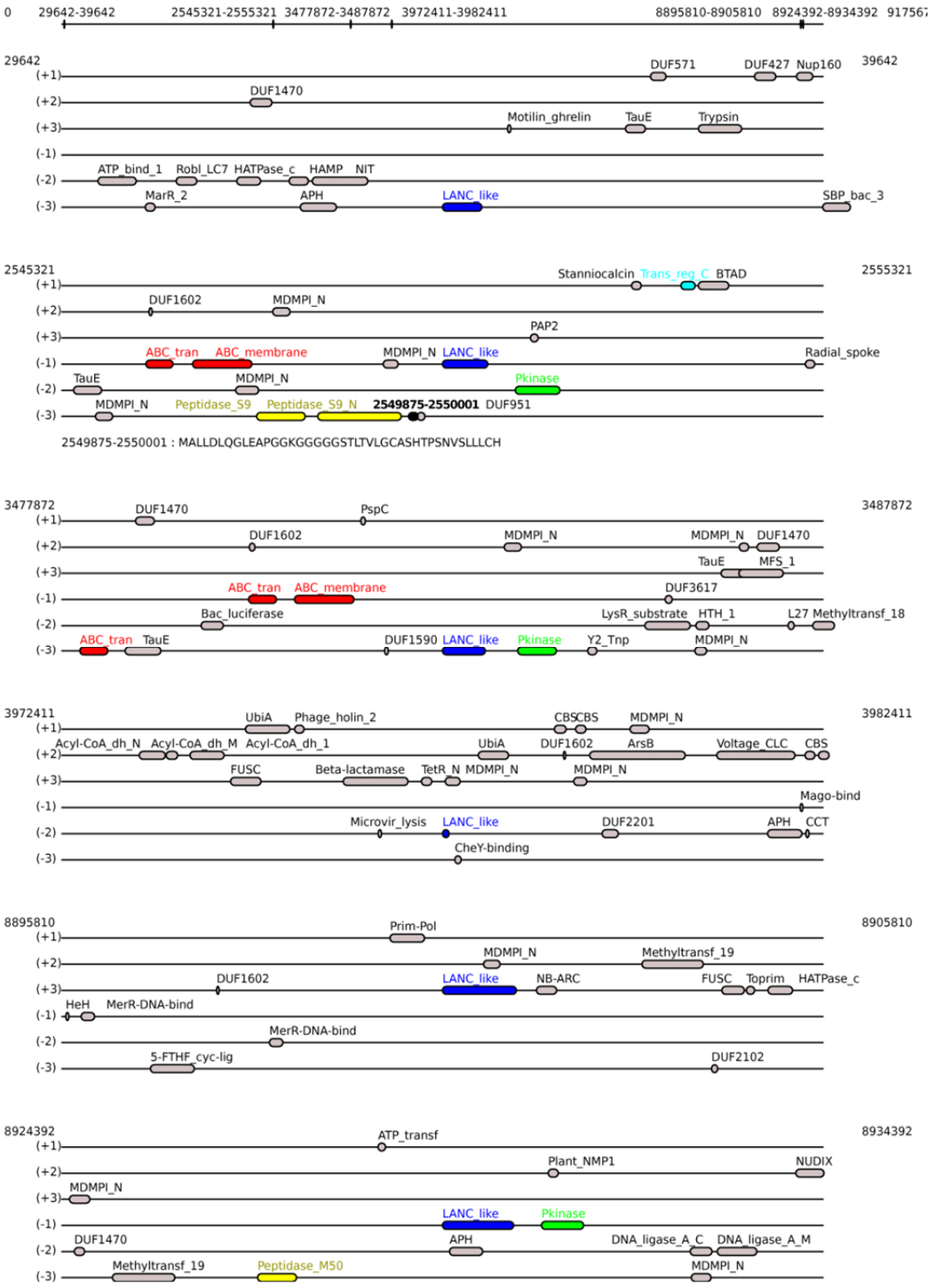
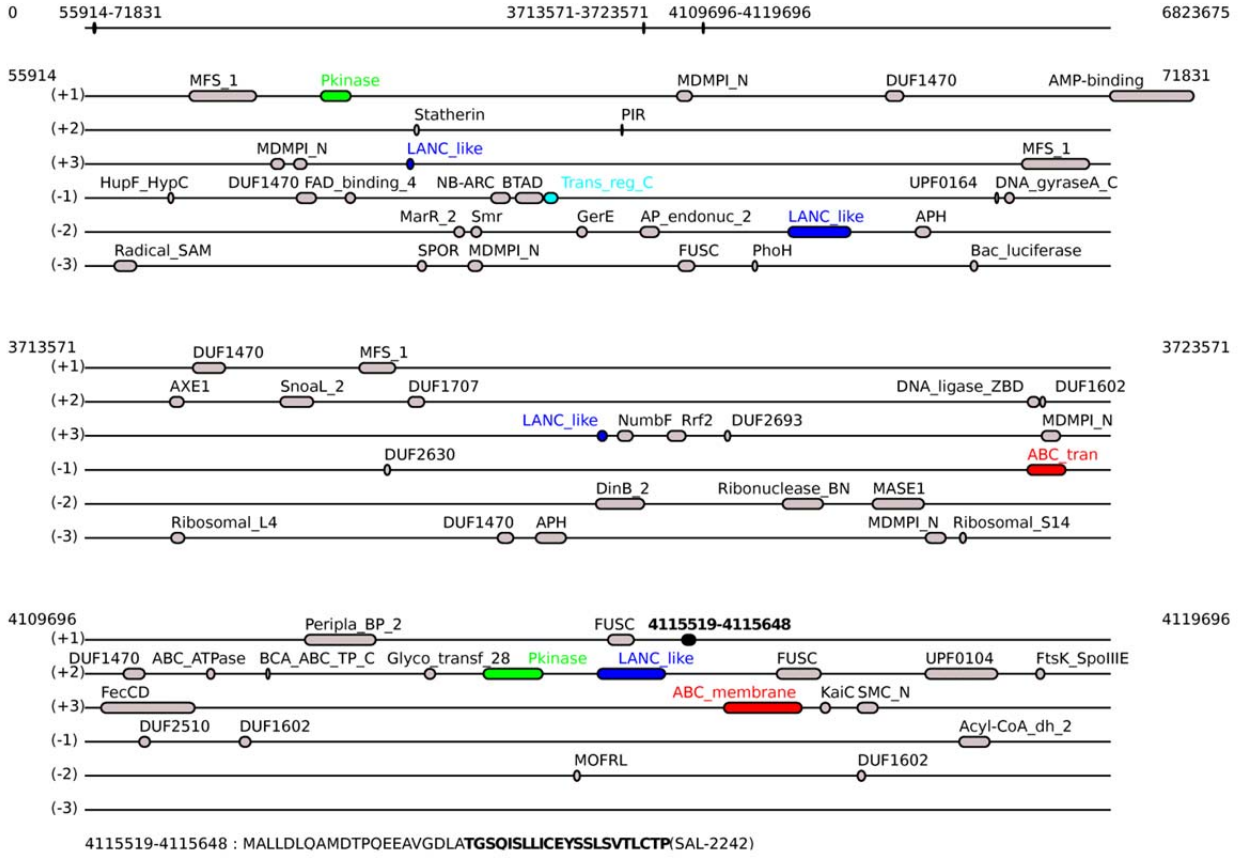


Figure S3. Lanthipeptide gene clusters predicted by RiPPquest in the genome of (C) *Streptomyces* sp. AA4. The figure shows all Pfam domains discovered in a 10,000 bp window centered at LANC_like domain, and predicted lanthipeptide precursor ORFs.

D – *Streptomyces albus* J1074



E – *Streptomyces sviceus* ATCC 29083

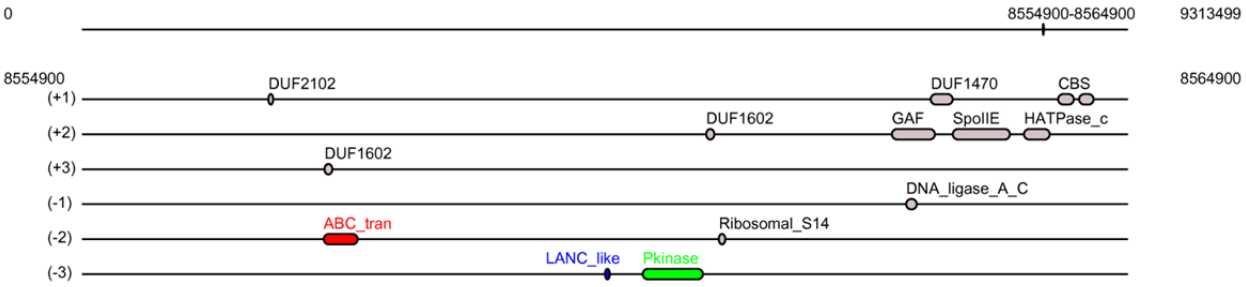
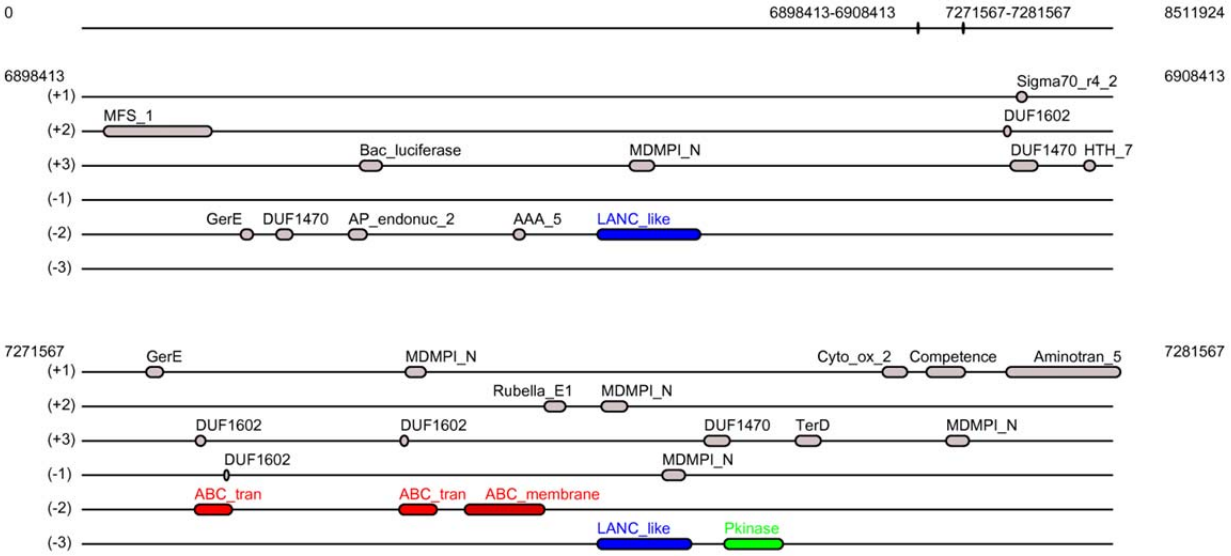


Figure S3. Lanthipeptide gene clusters predicted by RiPPquest in the genome of (D) *Streptomyces albus* J1074, (E) *Streptomyces sviceus* ATCC 29083. The figure shows all Pfam domains discovered in a 10,000 bp window centered at LANC_like domain, and predicted lanthipeptide precursor ORFs.

F - *Streptomyces ghanaensis* ATCC 14672



G - *Streptomyces pristinispiralis* ATCC 25486

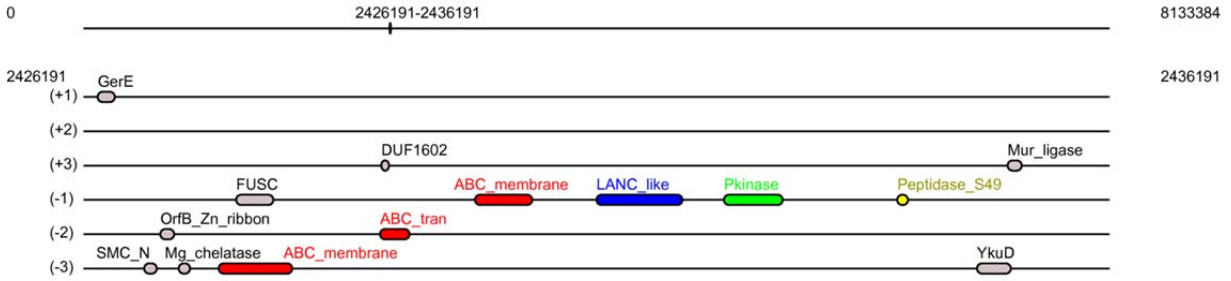


Figure S3. Lanthipeptide gene clusters predicted by RiPPquest in the genome of (F) *Streptomyces ghanaensis* ATCC 14672, (G) *Streptomyces griseoflavus* Tü4000. The figure shows all Pfam domains discovered in a 10,000 bp window centered at LANC_like domain, and predicted lanthipeptide precursor ORFs.

H - *Streptomyces griseoflavus* Tü4000

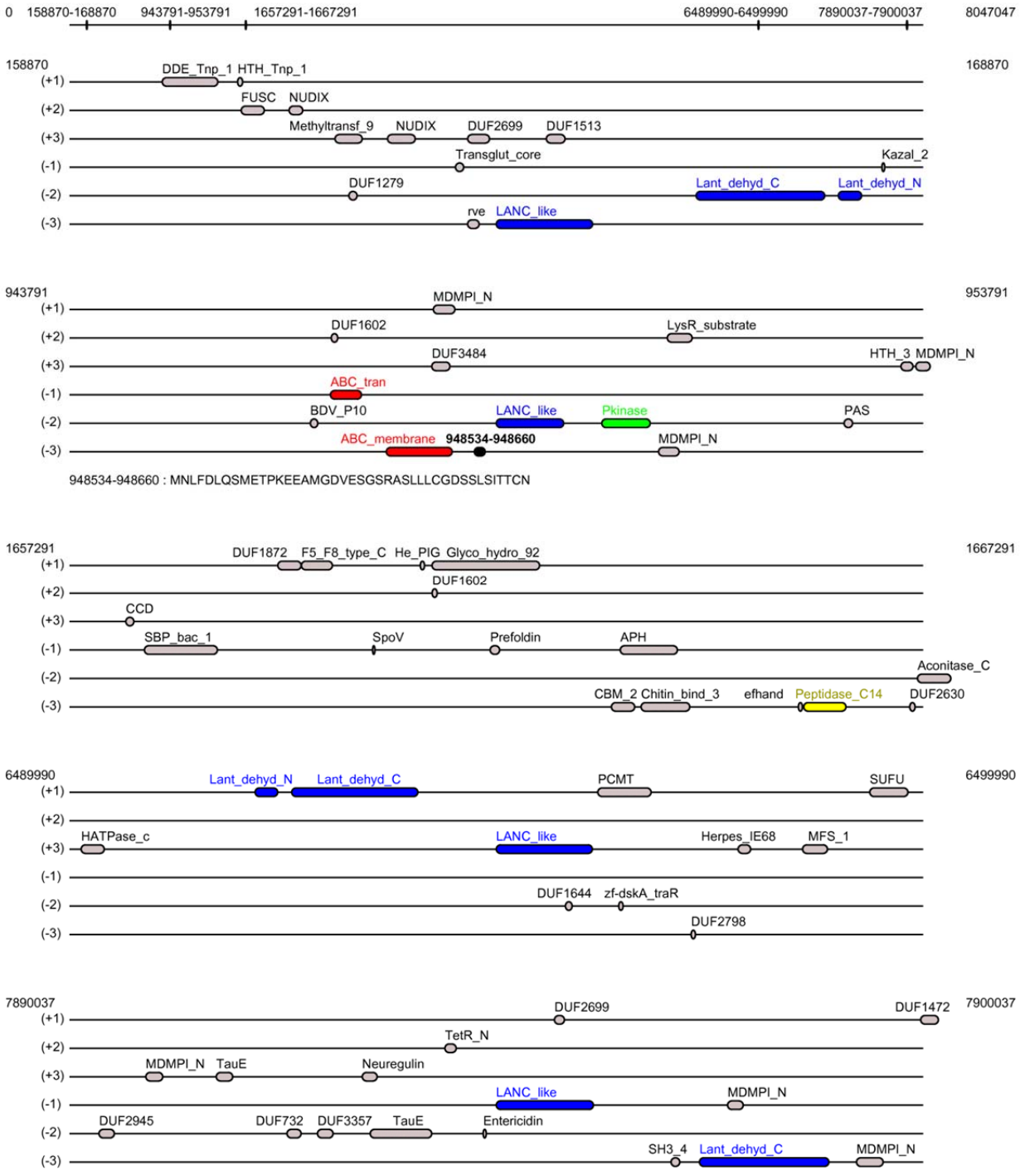


Figure S3. Lanthipeptide gene clusters predicted by RiPPquest in the genome of (H) *Streptomyces griseoflavus* Tü4000. The figure shows all Pfam domains discovered in a 10,000 bp window centered at LANC_like domain, and predicted lanthipeptide precursor ORFs.

I - *Streptomyces griseus* IFO 13350

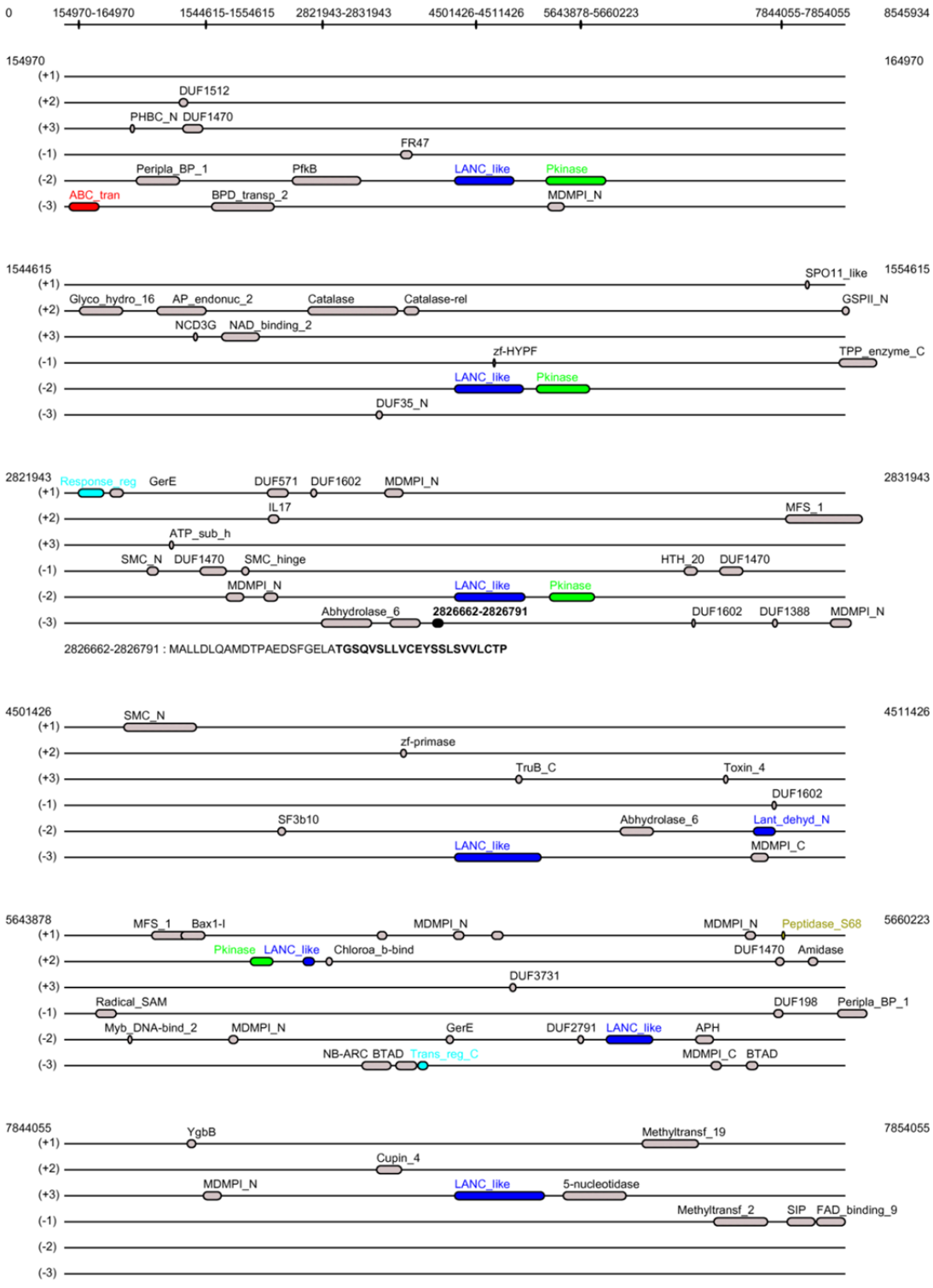
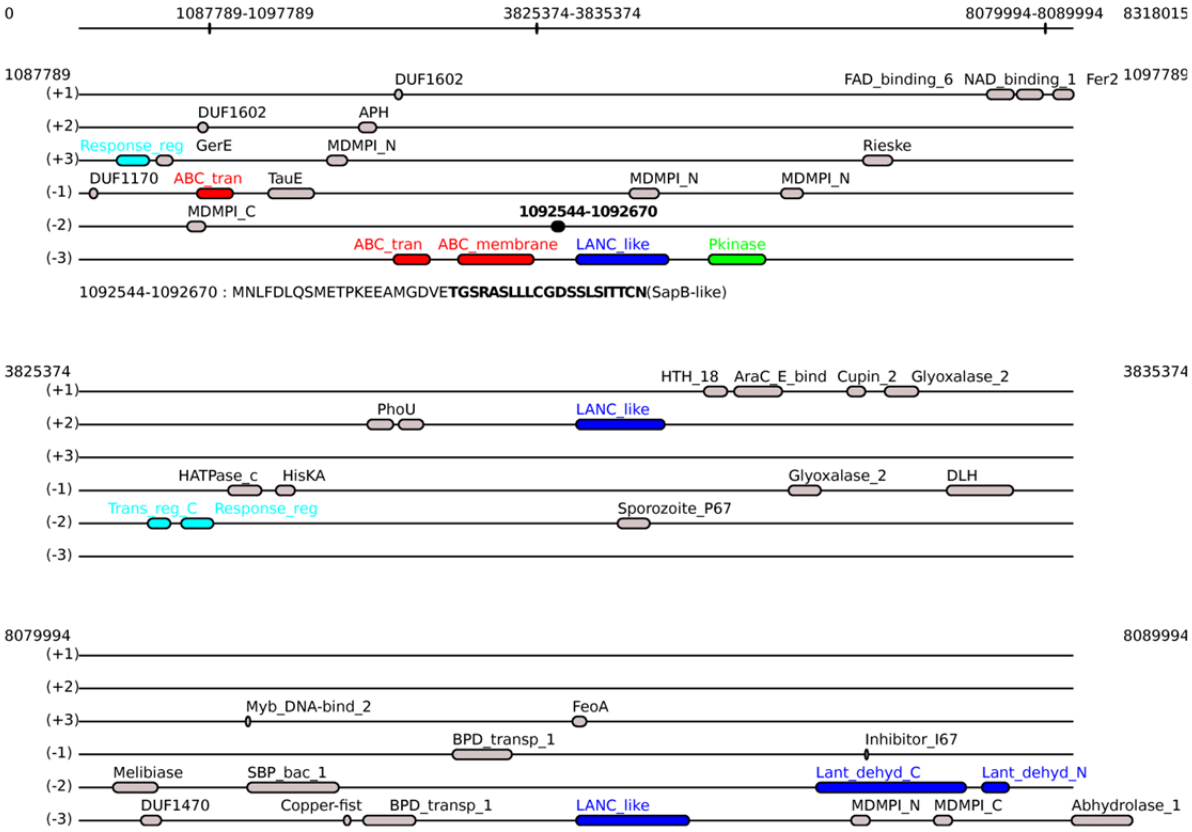


Figure S3. Lanthipeptide gene clusters predicted by RiPPquest in the genome of (I) *Streptomyces griseus* IFO 13350. The figure shows all Pfam domains discovered in a 10,000 bp window centered at LANC_like domain, and predicted lanthipeptide precursor ORFs.

J - *Streptomyces lividans* TK24



K - *Streptomyces* sp. E14

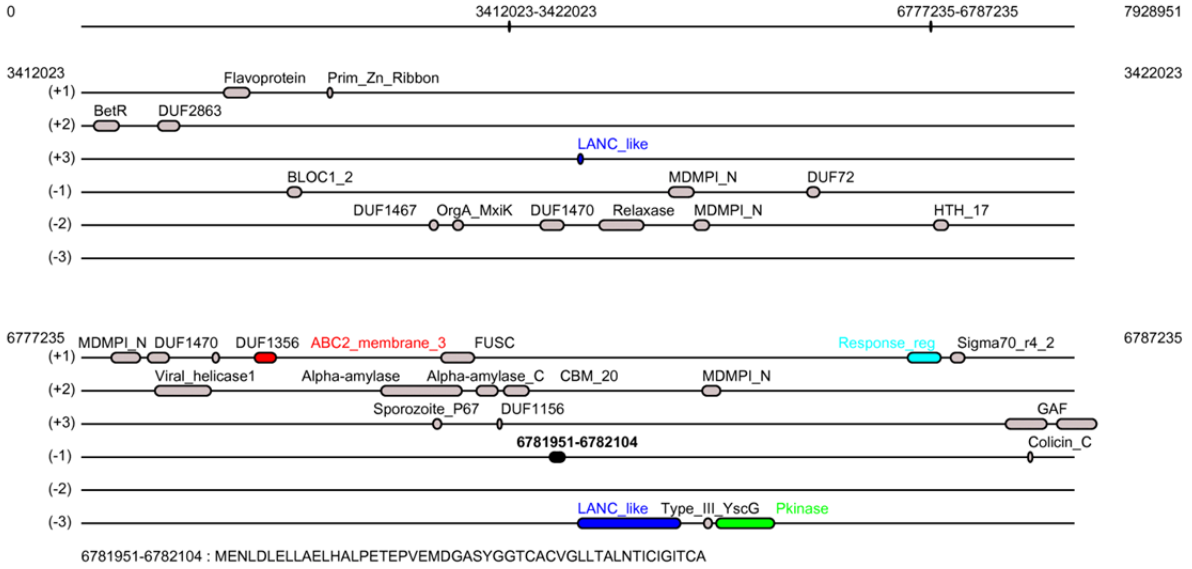
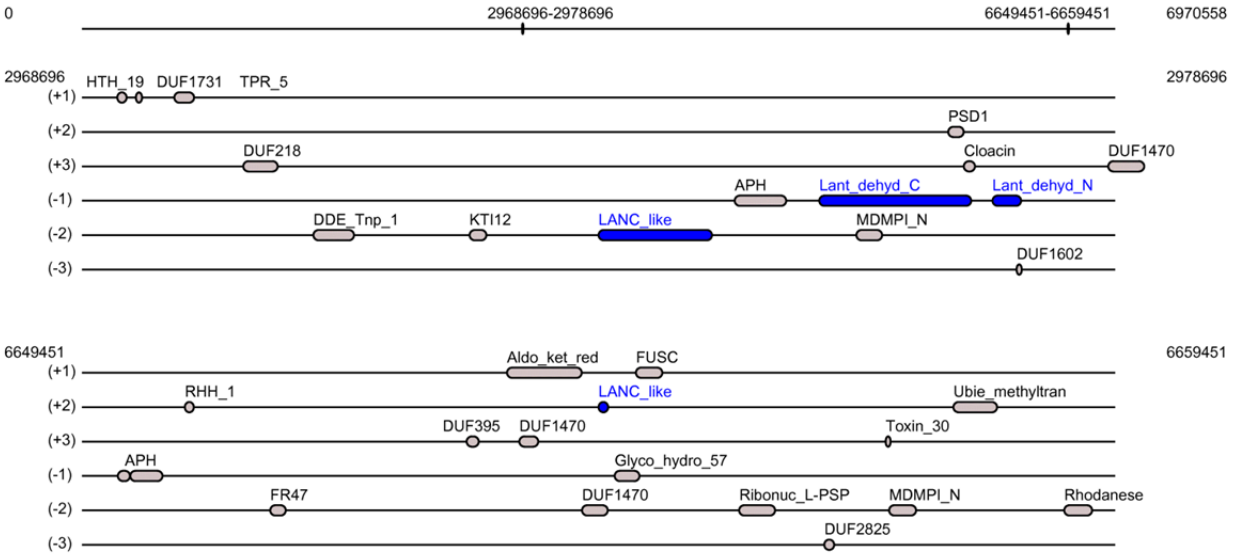


Figure S3. Lanthipeptide gene clusters predicted by RiPPquest in the genome of (J) *Streptomyces lividans* TK24**,** (K) *Streptomyces* sp. E14. The figure shows all Pfam domains discovered in a 10,000 bp window centered at LANC_like domain, and predicted lanthipeptide precursor ORFs.

L - *Streptomyces* sp. SPB74



M - *Streptomyces* sp. SPB78

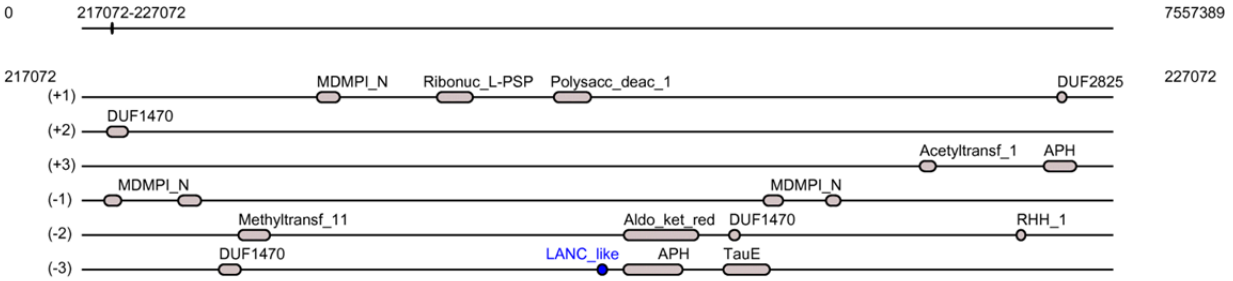


Figure S3. Lanthipeptide gene clusters predicted by RiPPquest in the genome of (L) *Streptomyces* sp. SPB74**,** (M) *Streptomyces* sp. SPB78*.* The figure shows all Pfam domains discovered in a 10,000 bp window centered at LANC_like domain, and predicted lanthipeptide precursor ORFs.

N - *Streptomyces viridochromogenes* DSM 40736

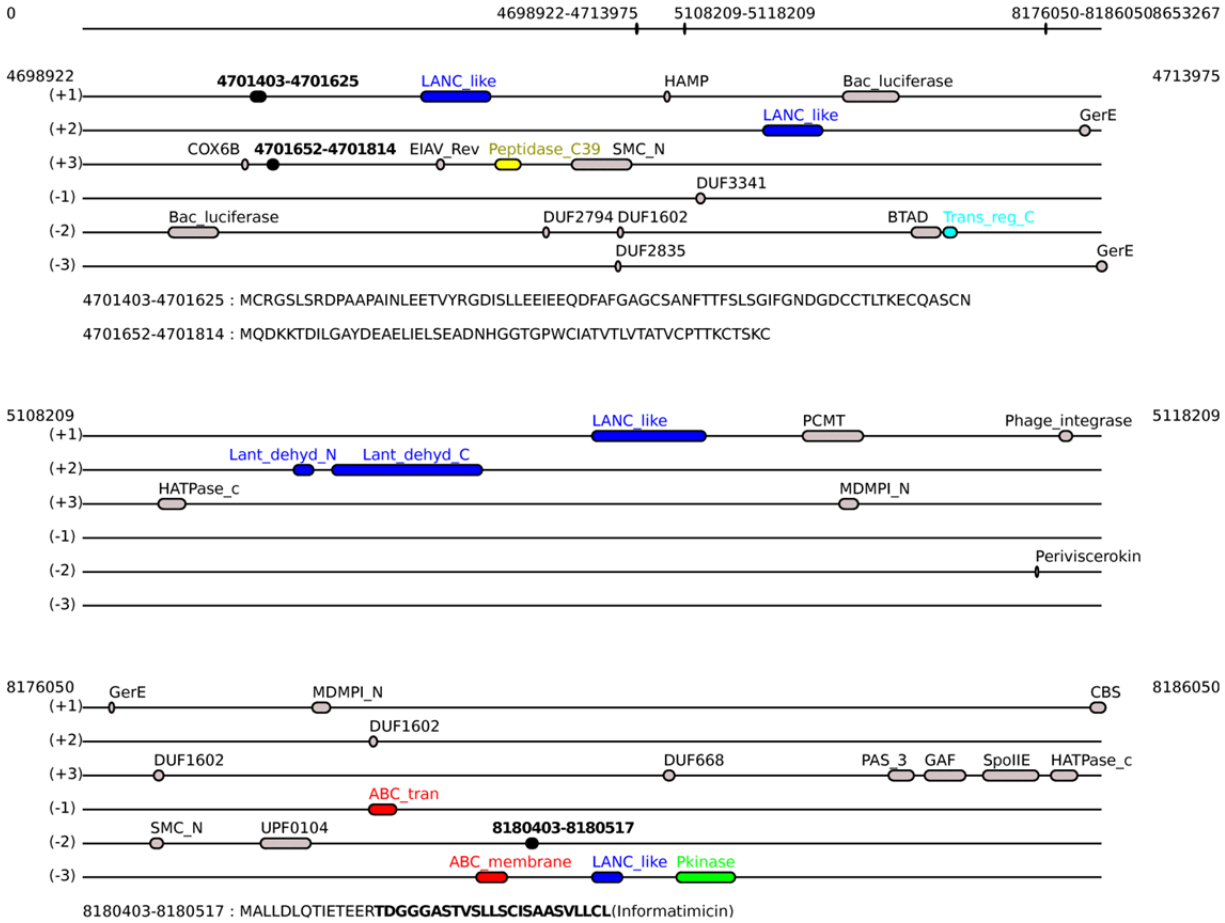


Figure S3. Lanthipeptide gene clusters predicted by RiPPquest in the genome of (N) *Streptomyces viridochromogenes* DSM 40736. The figure shows all Pfam domains discovered in a 10,000 bp window centered at LANC_like domain, and predicted lanthipeptide precursor ORFs.

O - *Streptomyces* sp. Mg1

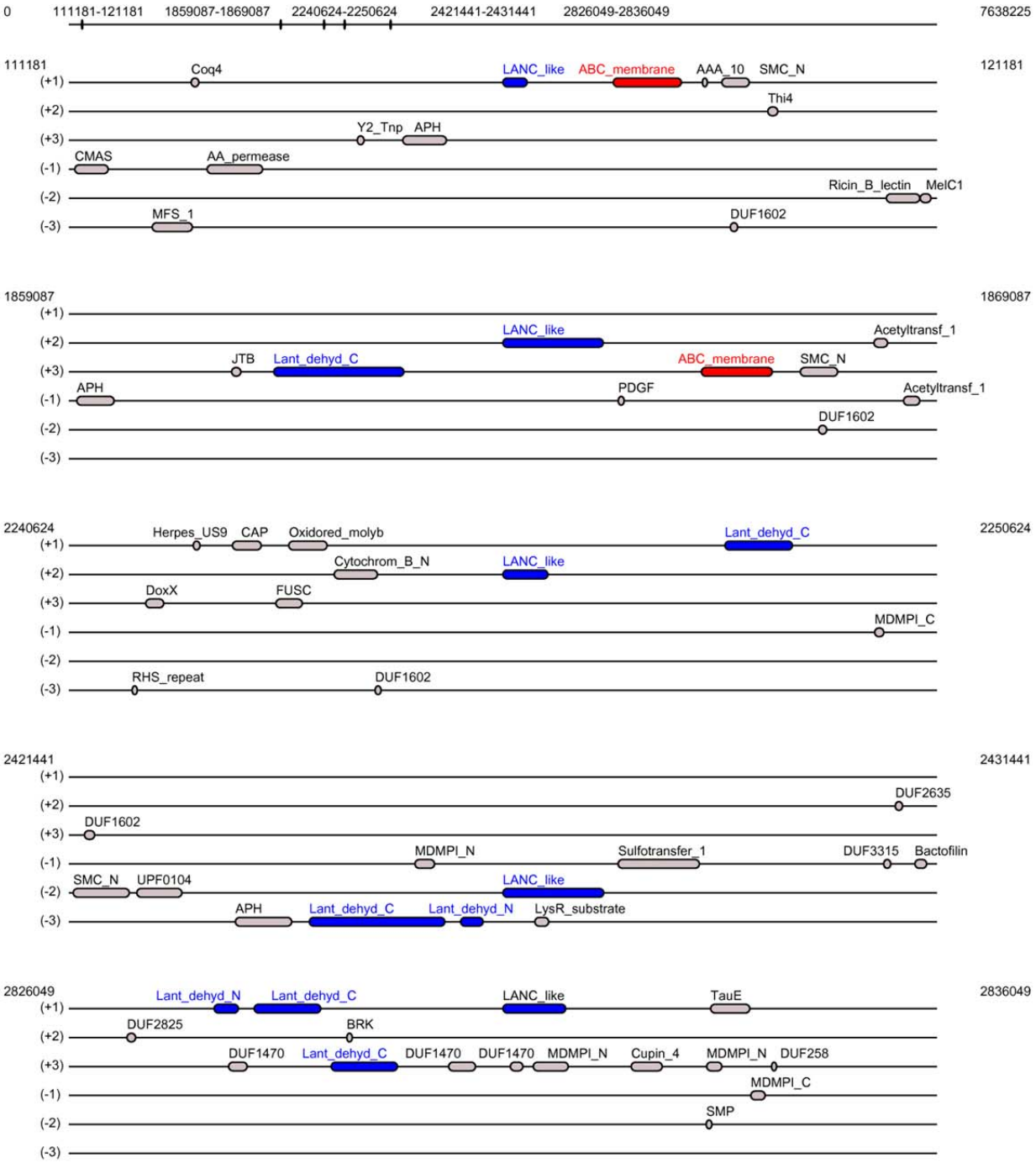
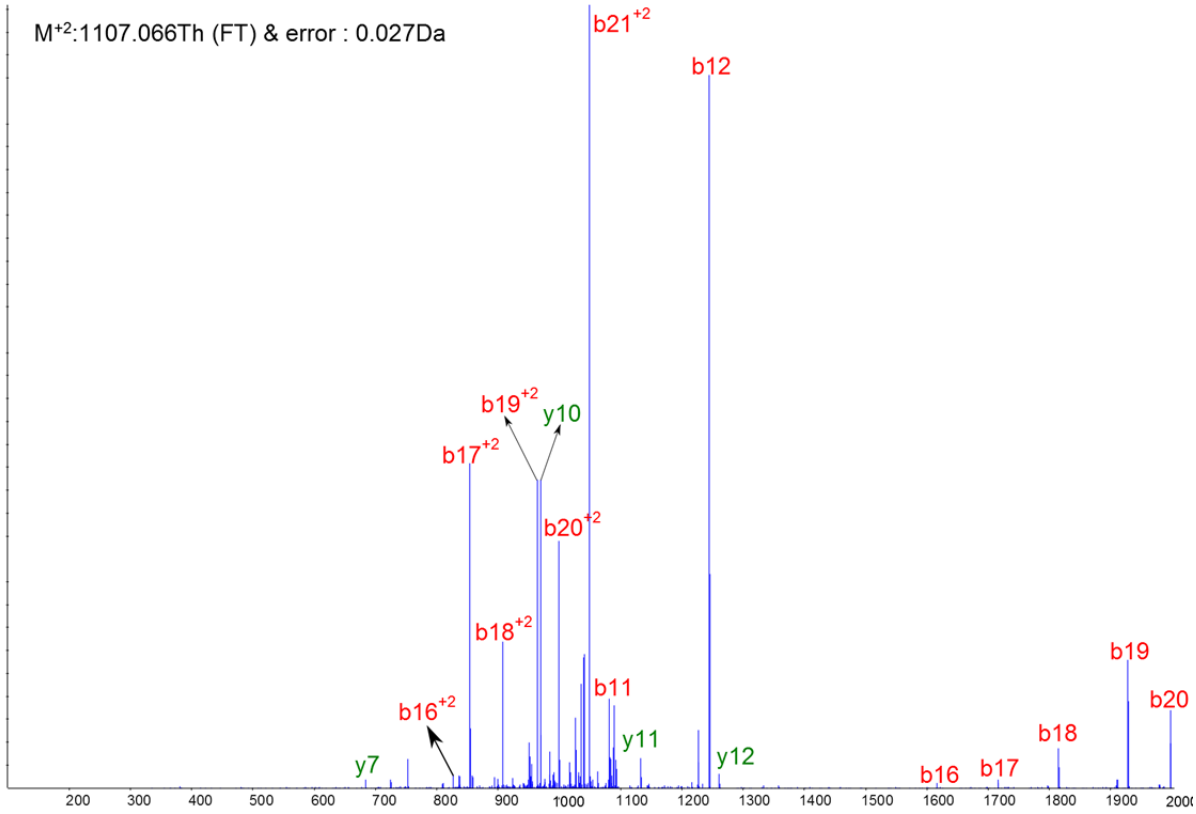


Figure S3. Lanthipeptide gene clusters predicted by RiPPquest in the genome of (O) *Streptomyces* sp. Mg1. The figure shows all Pfam domains discovered in a 10,000 bp window centered at LANC_like domain, and predicted lanthipeptide precursor ORFs.

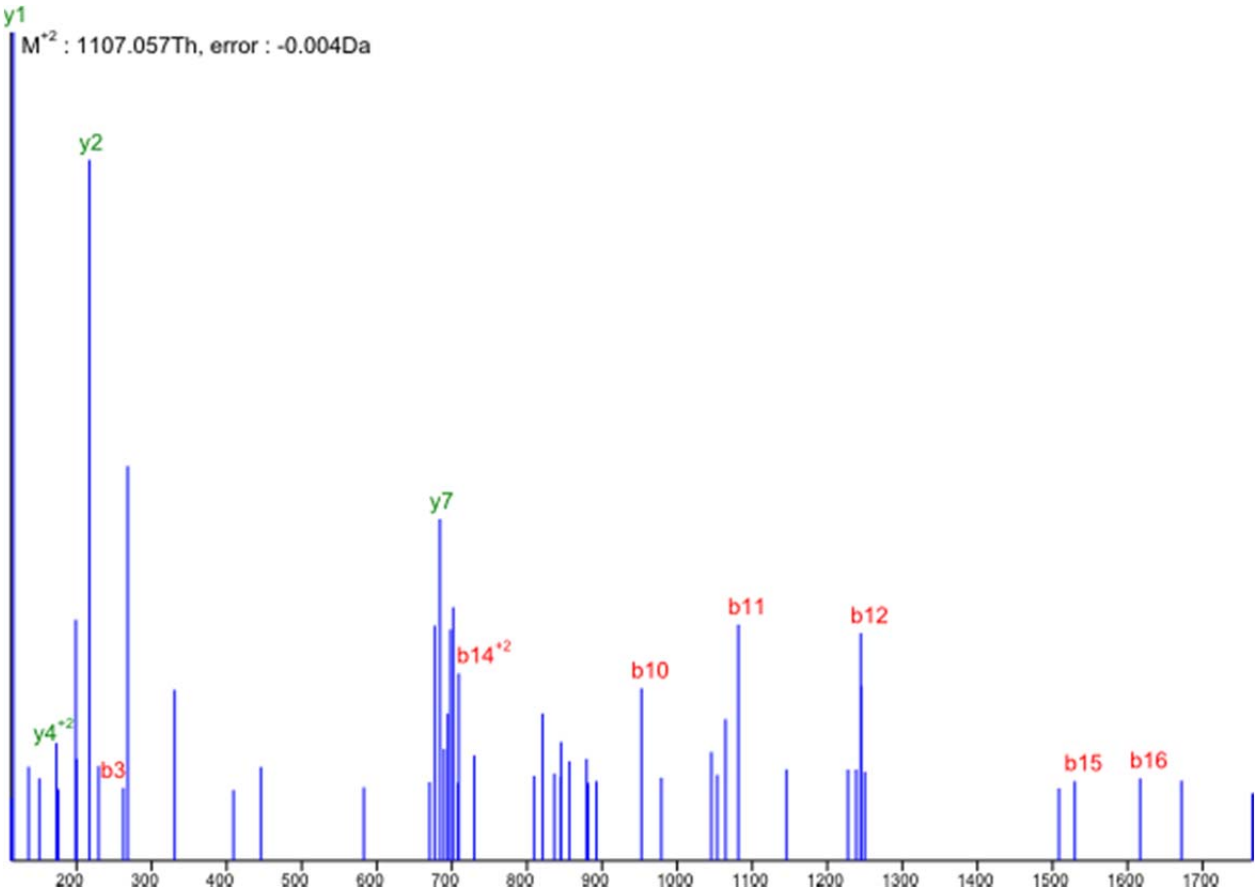
P - *Streptomyces coelicolor* A3(2)

Figure S3. Lanthipeptide gene clusters predicted by RiPPquest in the genome of (P) *Streptomyces coelicolor A3(2)*. The figure shows all Pfam domains discovered in a 10,000 bp window centered at LANC_like domain, and predicted lanthipeptide precursor ORFs.



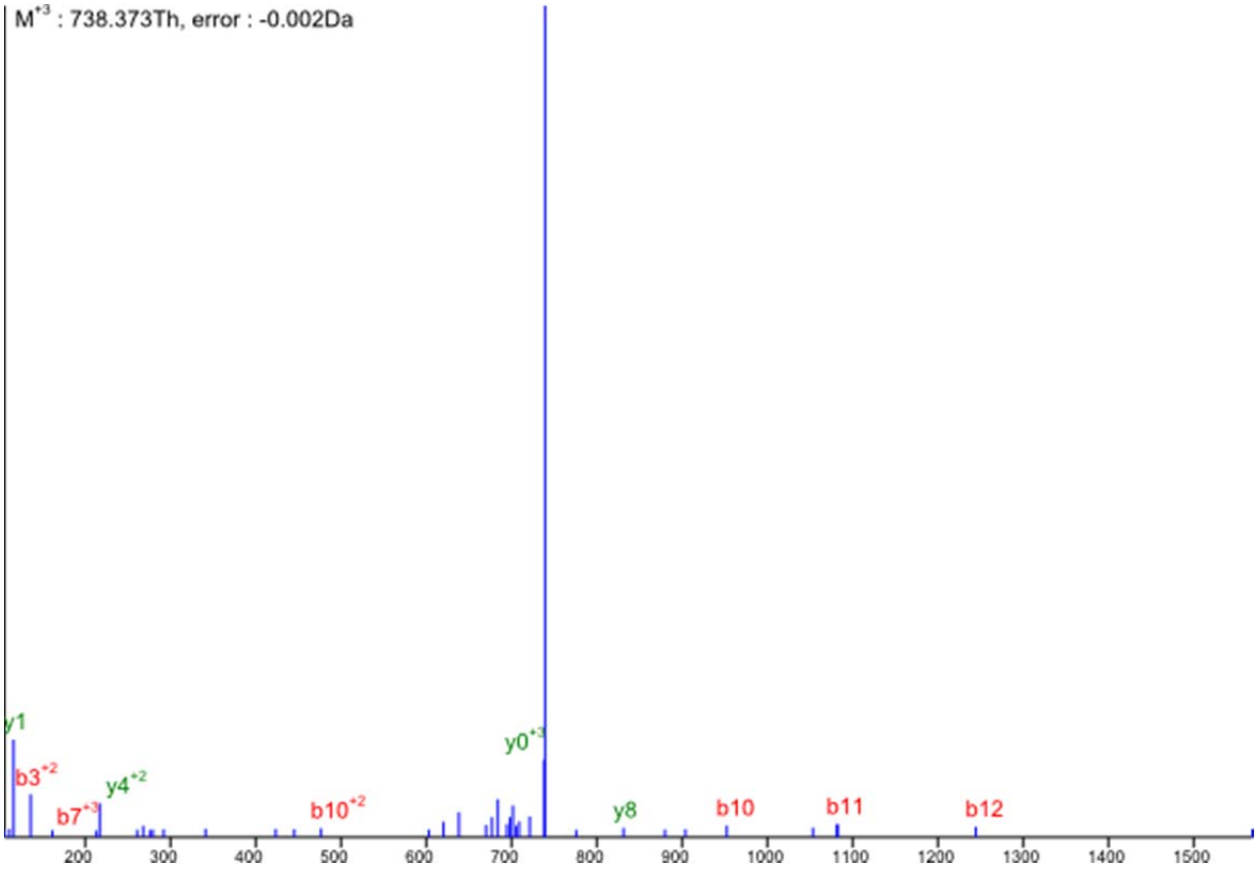
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
IT	684.374	1	y7	0.018
IT	808.801	2	b16	-0.11
IT	858.316	2	b17	-0.119
IT	907.975	2	b18	0.014
IT	964.52	2	b19	0.031
IT	969.291	1	y10	-0.206
IT	999.055	2	b20	0.049
IT	1049.554	2	b21	0.022
IT	1081.304	1	b11	-0.243
IT	1132.227	1	y11	-0.241
IT	1244.344	1	b12	-0.284
IT	1261.39	1	y12	-0.253
IT	1616.468	1	b16	-0.347
IT	1715.515	1	b17	-0.348
IT	1814.506	1	b18	-0.407
IT	1927.638	1	b19	-0.332
IT	1997.123	1	b20	0.117

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. CID spectrum of TGSQVSLLVCEYSSLSVVLCTP (+2) with annotations from *S. roseosporus* NRRL 15998 (4 Dehyd).



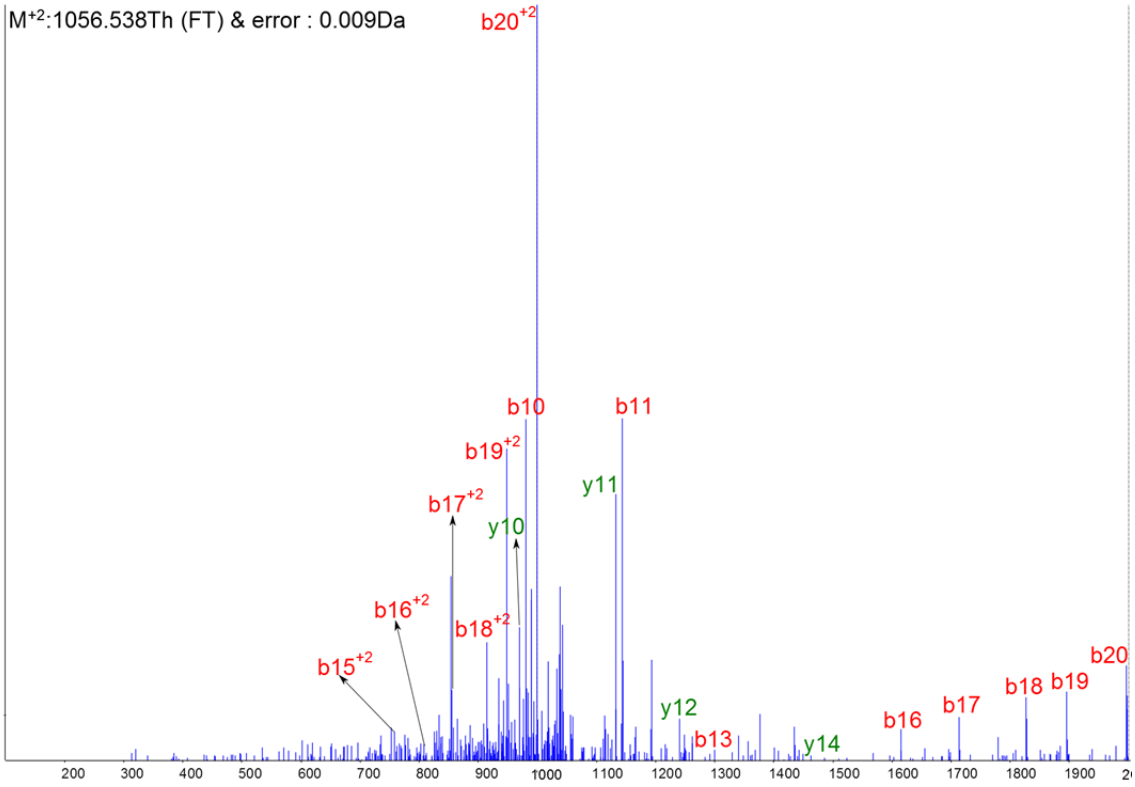
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
orbitrap	116.071	1	y1	0
orbitrap	200.11	2	y4	0.029
orbitrap	217.122	1	y2	-0.002
orbitrap	262.138	1	b3	0.2
orbitrap	684.355	1	y7	-0.315
orbitrap	708.861	2	b14	0.166
orbitrap	952.483	1	b10	0.018
orbitrap	1081.55	1	b11	-0.006
orbitrap	1244.63	1	b12	-0.028
orbitrap	1529.77	1	b15	-0.031
orbitrap	1616.82	1	b16	-0.081

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. HCD spectrum of TGSQVSLLVCEYSSLSVVLCTP (+2) with annotations from *S. roseosporus* NRRL 15998 (4 Dehyd).



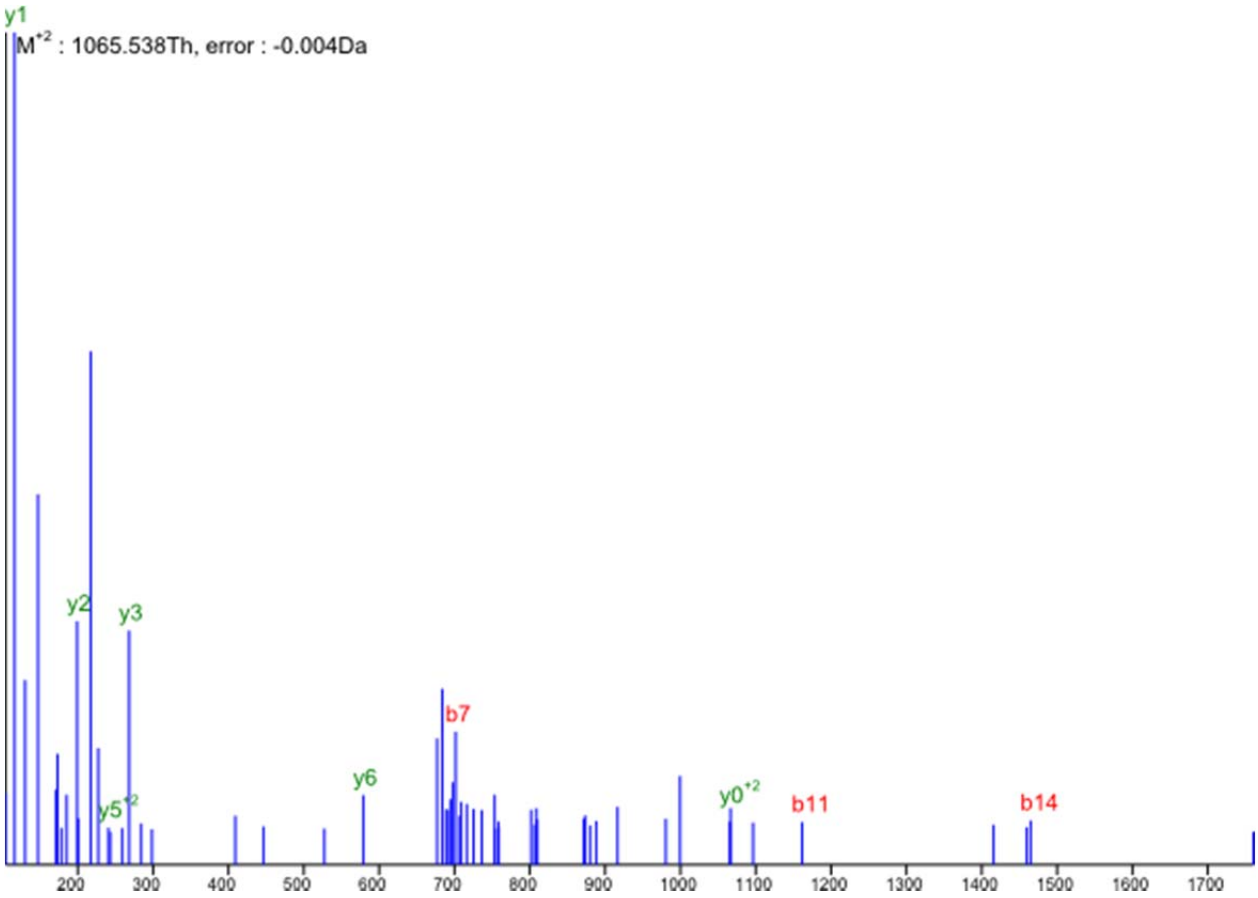
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
orbitrap	105.56	2	b3	0.077
orbitrap	116.071	1	y1	0
orbitrap	213.113	3	b7	0.079
orbitrap	217.118	2	y4	0
orbitrap	476.745	2	b10	-0.005
orbitrap	738.378	3	y0	0.094
orbitrap	831.429	1	y8	-0.006
orbitrap	952.483	1	b10	0.018
orbitrap	1081.55	1	b11	-0.004
orbitrap	1244.63	1	b12	-0.041

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. HCD spectrum of GSQVSLLVCEYSSLSVVLCTP (+3) with annotations from *S. roseosporus* NRRL 15998 (4 Dehyd).



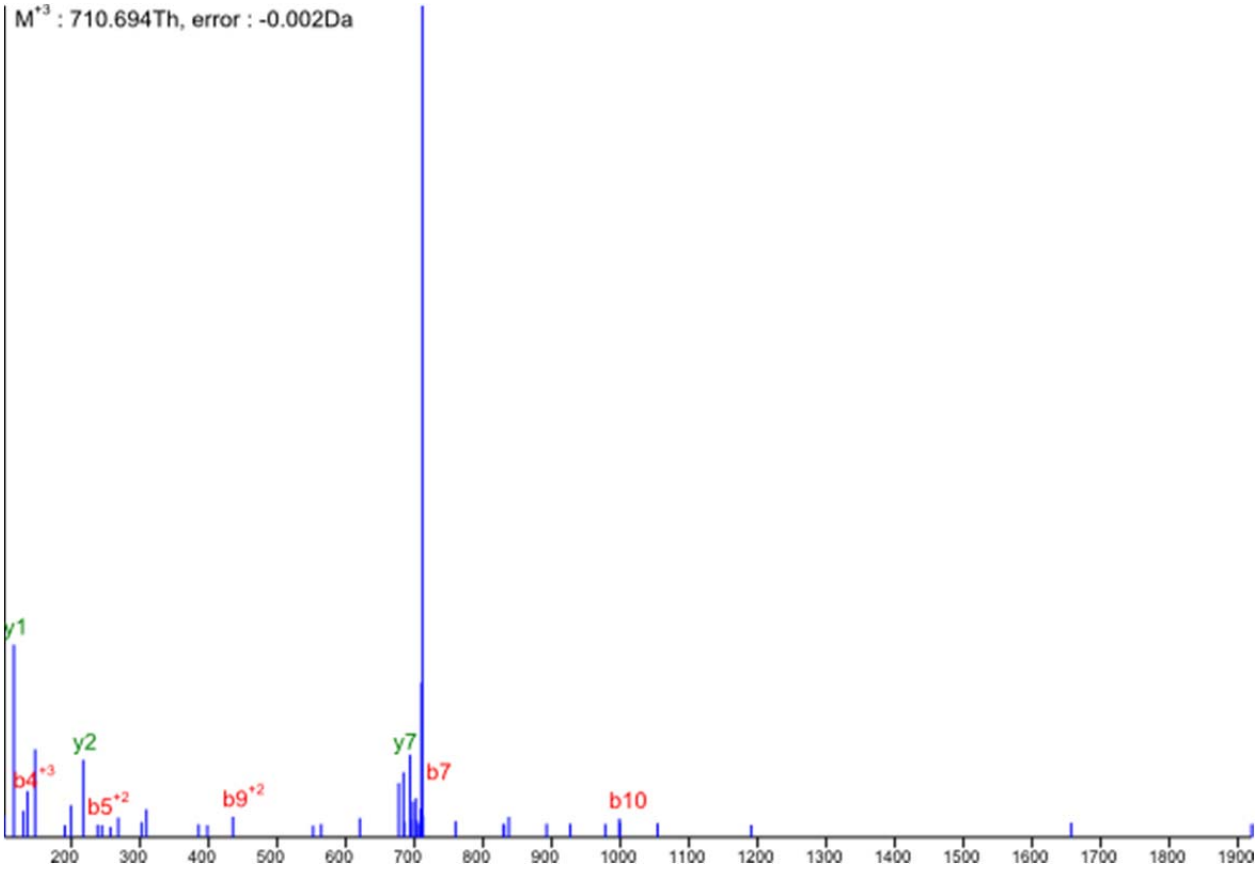
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
IT	758.359	2	b15	0.026
IT	808.125	2	b16	0.214
IT	857.547	2	b17	0.112
IT	914.118	2	b18	0.154
IT	948.306	2	b19	0.174
IT	969.093	1	y10	-0.404
IT	980.306	1	b10	-0.19
IT	998.957	2	b20	0.048
IT	1132.319	1	y11	-0.26
IT	1143.349	1	b11	-0.229
IT	1261.442	1	y12	-0.201
IT	1299.672	1	b13	0.015
IT	1463.448	1	y14	-0.296
IT	1614.416	1	b16	-0.398
IT	1713.772	1	b17	-0.091
IT	1826.578	1	b18	-0.341
IT	1895.705	1	b19	-0.249
IT	1996.7	1	b20	0.362

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. CID spectrum of GSQVSLLVCEYSSLSVVLCTP (+2) with annotations from *S. roseosporus* NRRL 15998 (4 Dehyd).



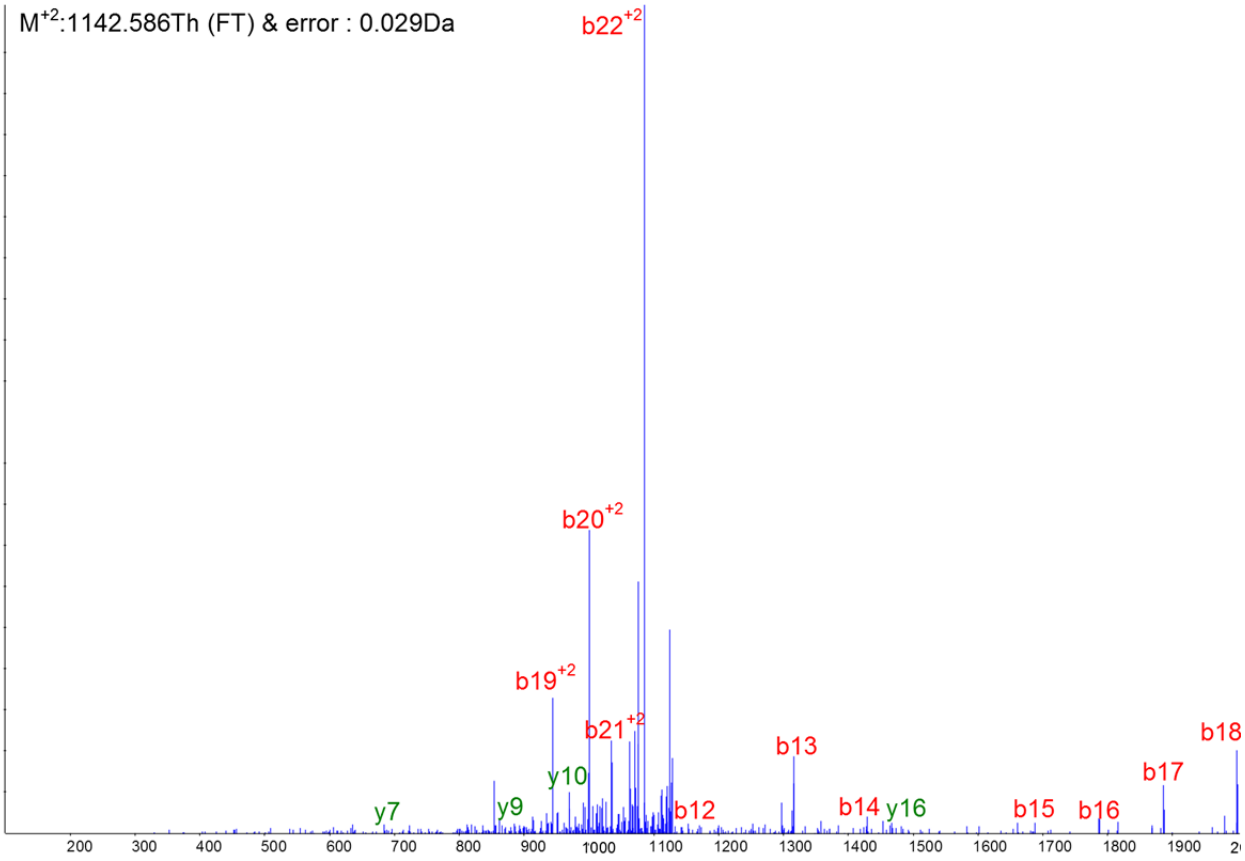
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
orbitrap	116.071	1	y1	0
orbitrap	199.113	1	y2	0.028
orbitrap	240.63	2	y5	0.075
orbitrap	268.147	1	y3	-0.017
orbitrap	579.303	1	y6	-0.051
orbitrap	701.357	1	b7	0.317
orbitrap	1065.54	2	y0	-0.029
orbitrap	1161.59	1	b11	-0.034
orbitrap	1464.74	1	b14	0.224

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. HCD spectrum of GSQVSLLVCEYSSLSVVLCTP (+2) with annotations from *S. roseosporus* NRRL 15998 (3 Dehyd).



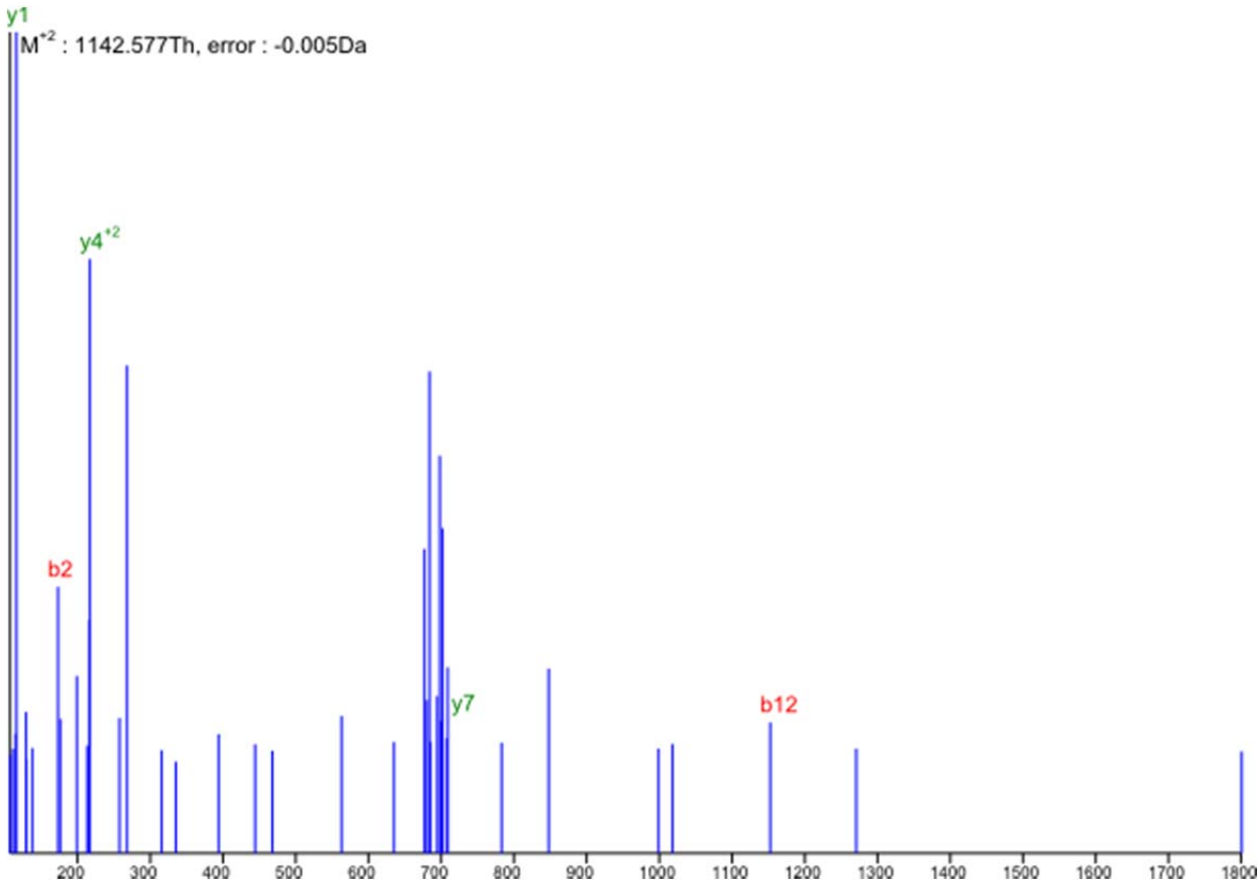
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
orbitrap	116.071	1	y1	0
orbitrap	130.072	3	b4	-0.021
orbitrap	217.122	1	y2	-0.002
orbitrap	238.126	2	b5	0.095
orbitrap	435.224	2	b9	0.005
orbitrap	684.355	1	y7	-0.384
orbitrap	701.357	1	b7	0.339
orbitrap	998.506	1	b10	0

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. HCD spectrum of GSQVSLLVCEYSSLSVVLCTP (+3) with annotations from *S. roseosporus* NRRL 15998 (3 Dehyd).



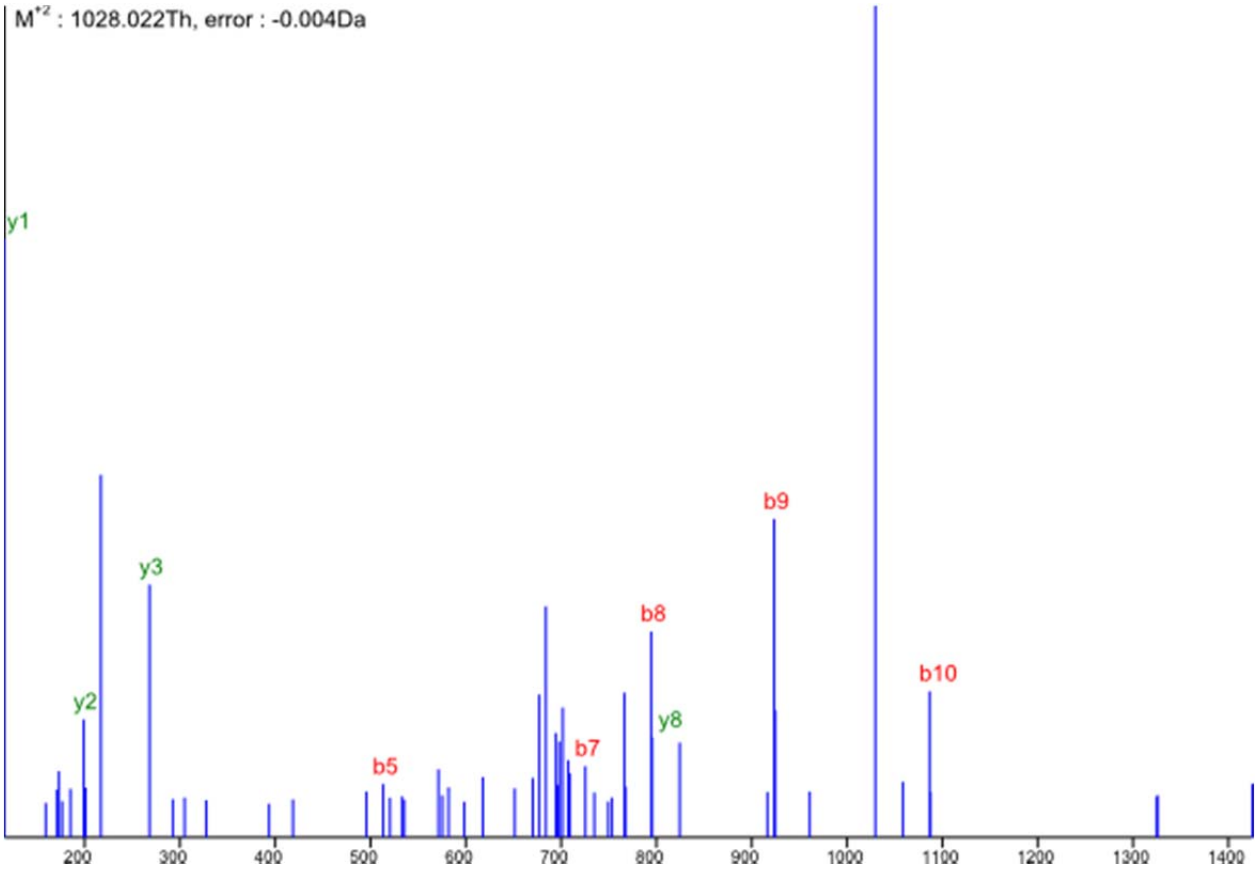
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
IT	684.342	1	y7	-0.012
IT	866.195	1	y9	-0.25
IT	943.586	2	b19	0.108
IT	969.098	1	y10	-0.399
IT	1000.114	2	b20	0.107
IT	1034.469	2	b21	-0.054
IT	1085.049	2	b22	0
IT	1152.371	1	b12	-0.211
IT	1315.399	1	b13	-0.26
IT	1418.629	1	b14	-0.087
IT	1464.203	1	y16	0.458
IT	1687.354	1	b15	-0.496
IT	1786.749	1	b16	-0.15
IT	1885.873	1	b17	-0.076

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. CID spectrum of ATGSQVSLLVCEYSSLSVVLCTP (+2) with annotations from *S. roseosporus* NRRL 15998 (4 Dehyd).



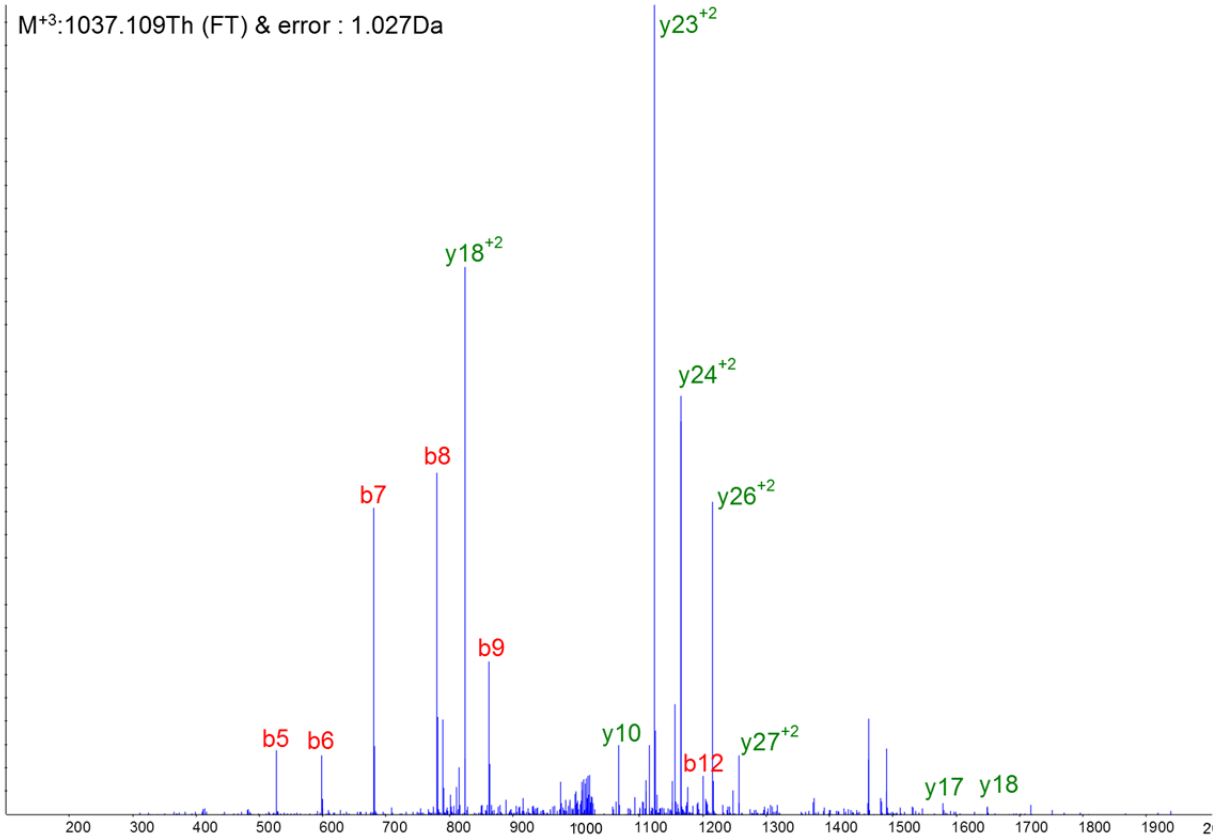
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
orbitrap	116.071	1	y1	0
orbitrap	173.093	1	b2	0.06
orbitrap	217.119	2	y4	0
orbitrap	700.364	1	y7	-0.456
orbitrap	1152.58	1	b12	-0.007

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. HCD spectrum of ATGSQVSLLVCEYSSLSVVLCTP (+2) with annotations from *S. roseosporus* NRRL 15998 (4 Dehyd).



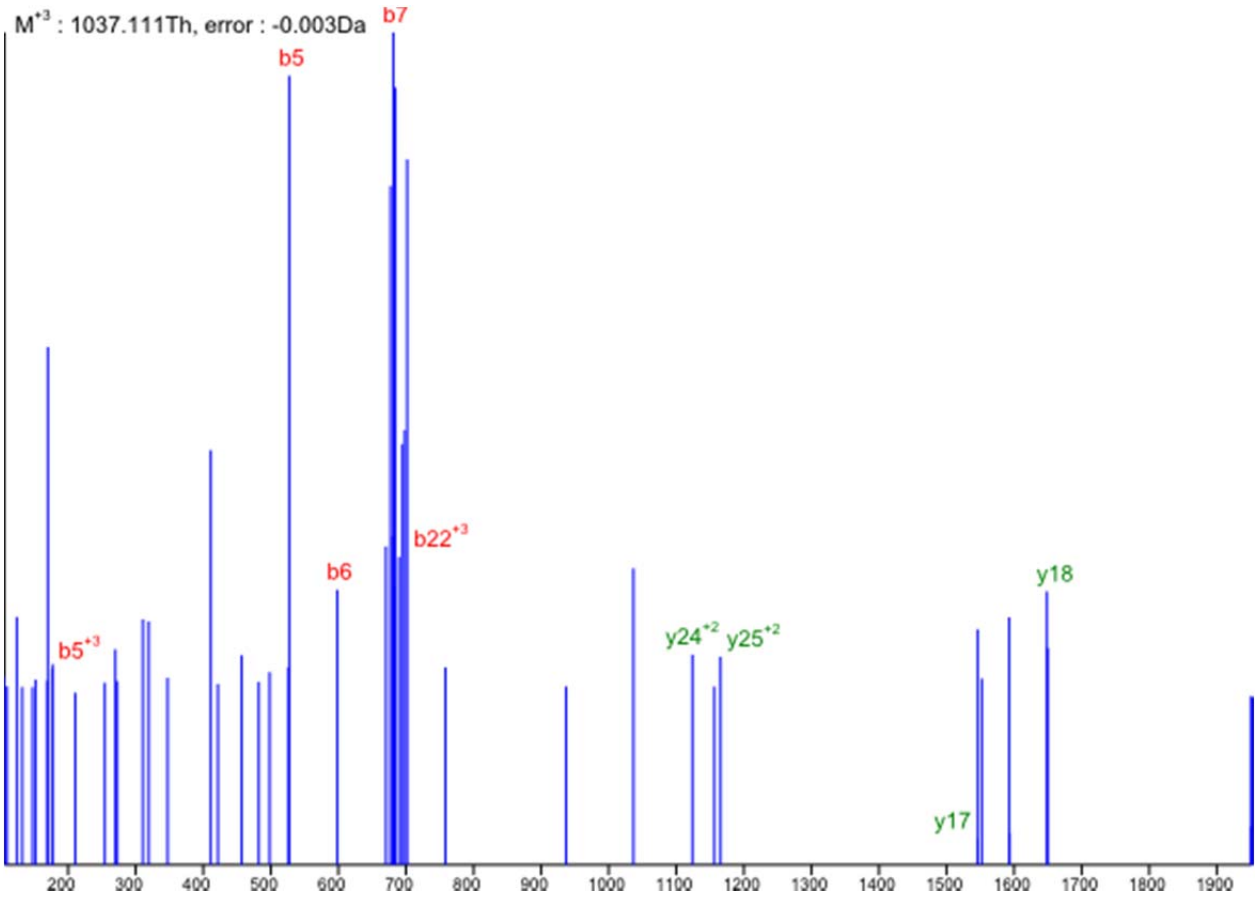
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
orbitrap	116.071	1	y1	-0.006
orbitrap	199.113	1	y2	-0.002
orbitrap	268.147	1	y3	-0.017
orbitrap	513.263	1	b5	-0.018
orbitrap	725.369	1	b7	0.033
orbitrap	794.404	1	b8	0.0156
orbitrap	795.411	1	y8	0.015
orbitrap	923.469	1	b9	-0.002
orbitrap	1086.55	1	b10	-0.013

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. HCD spectrum of SQVSLLVCEYSSLSVVLCTP (+2) with annotations from *S. roseosporus* NRRL 15998 (4 Dehyd).



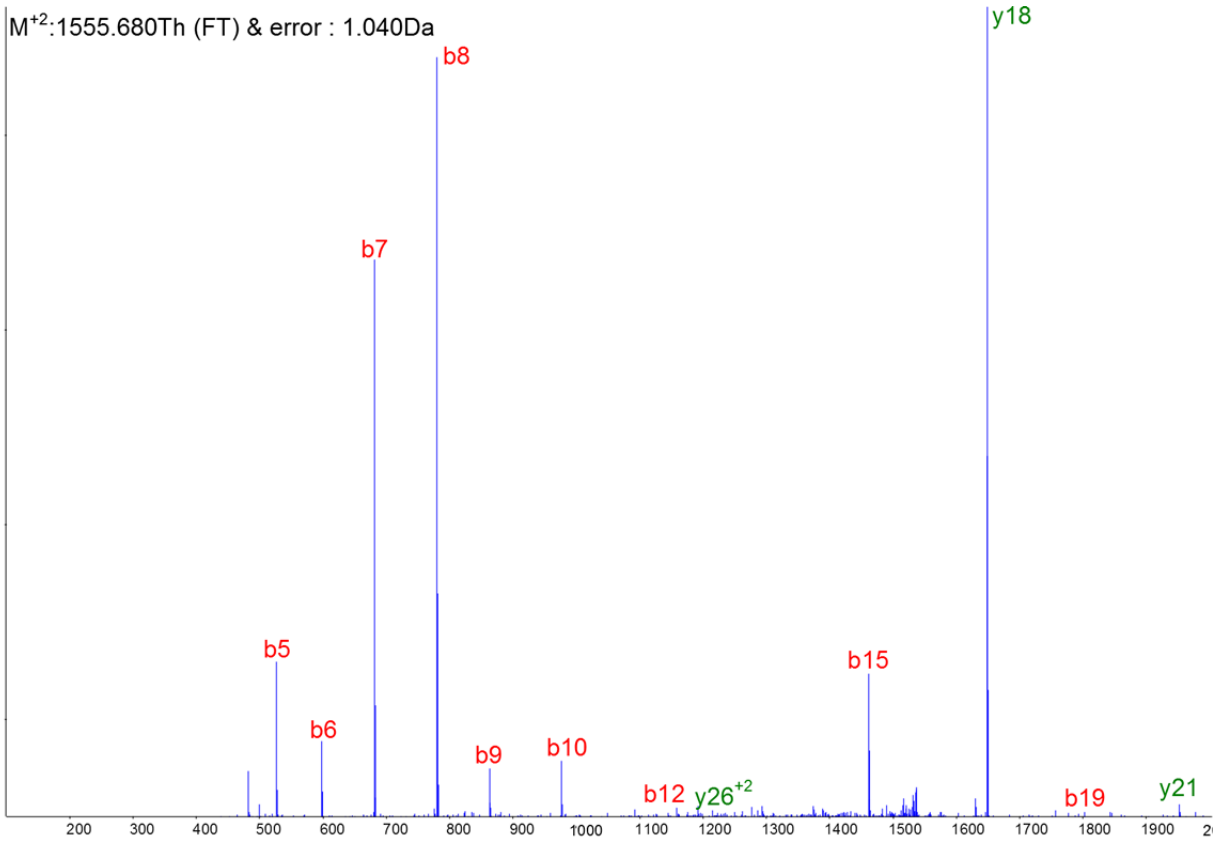
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
IT	527.052	1	b5	-0.218
IT	597.97	1	b6	-0.335
IT	681.103	1	b7	-0.243
IT	780.02	1	b8	-0.376
IT	825.07	2	y18	0.147
IT	862.963	1	b9	-0.474
IT	1020.574	1	y10	0.051
IT	1067.475	2	y23	-0.067
IT	1124.008	2	y24	-0.063
IT	1165.285	2	y25	0.432
IT	1215.15	2	y26	0.033
IT	1256.402	2	y27	-0.235
IT	1579.503	1	y17	-0.299
IT	1648.424	1	y18	-0.413

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. CID spectrum of TTWACATVTLTVTVCSPTGTLCGSCSMGTRGCC (+3) with annotations from *S. roseosporus* NRRL 15998 (9 Dehyd).



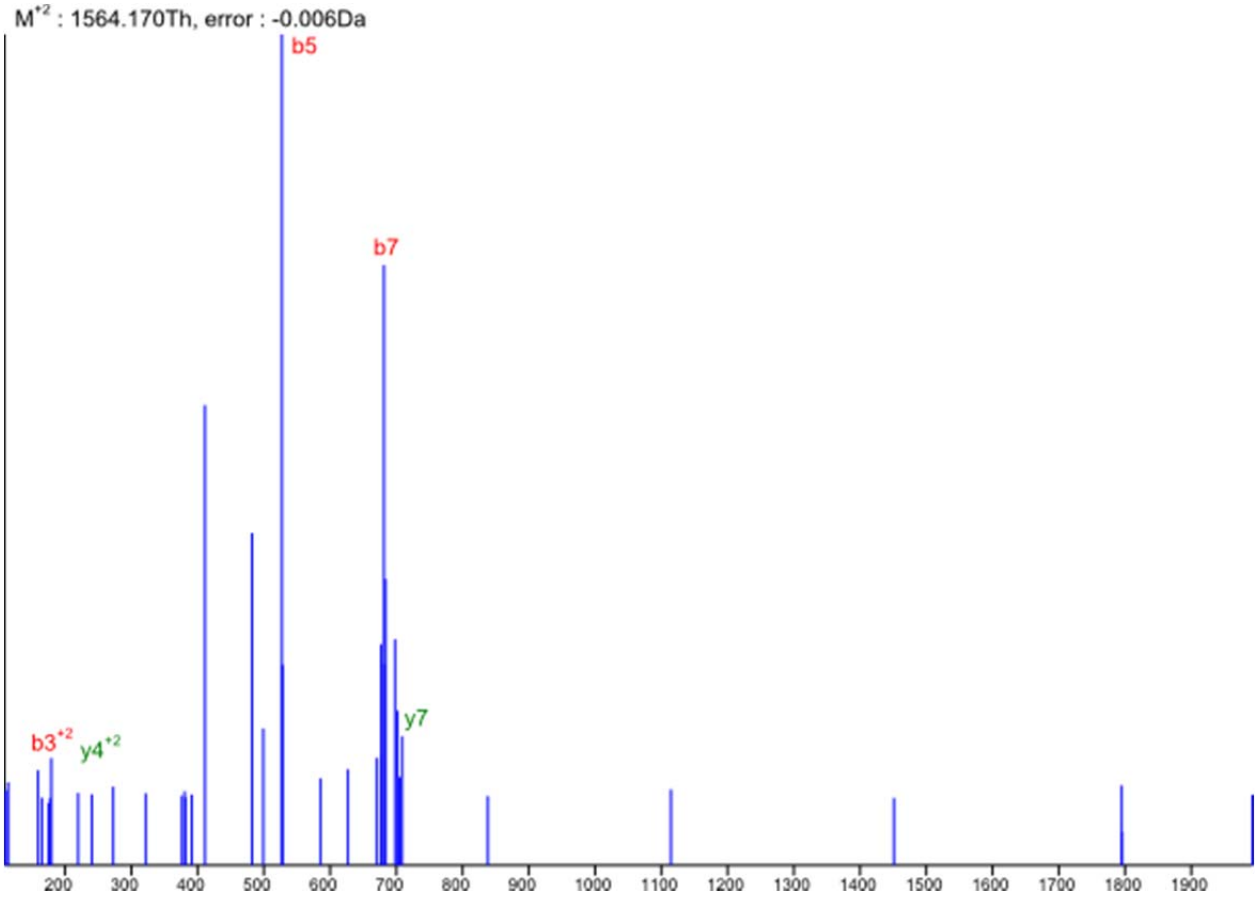
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
orbitrap	176.428	3	b5	-0.054
orbitrap	527.27	1	b5	-0.061
orbitrap	598.306	1	b6	-0.053
orbitrap	681.347	1	b7	-0.063
orbitrap	689.685	3	b22	0.148
orbitrap	1124.07	2	y24	-0.094
orbitrap	1165.59	2	y25	-0.119
orbitrap	1545.79	1	y17	0.457
orbitrap	1648.84	1	y18	-0.251

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. HCD spectrum of TTWACATVTLTVTVCSPTGTLCGSCSMGTRGCC (+3) with annotations from *S. roseosporus* NRRL 15998 (9 Dehyd).



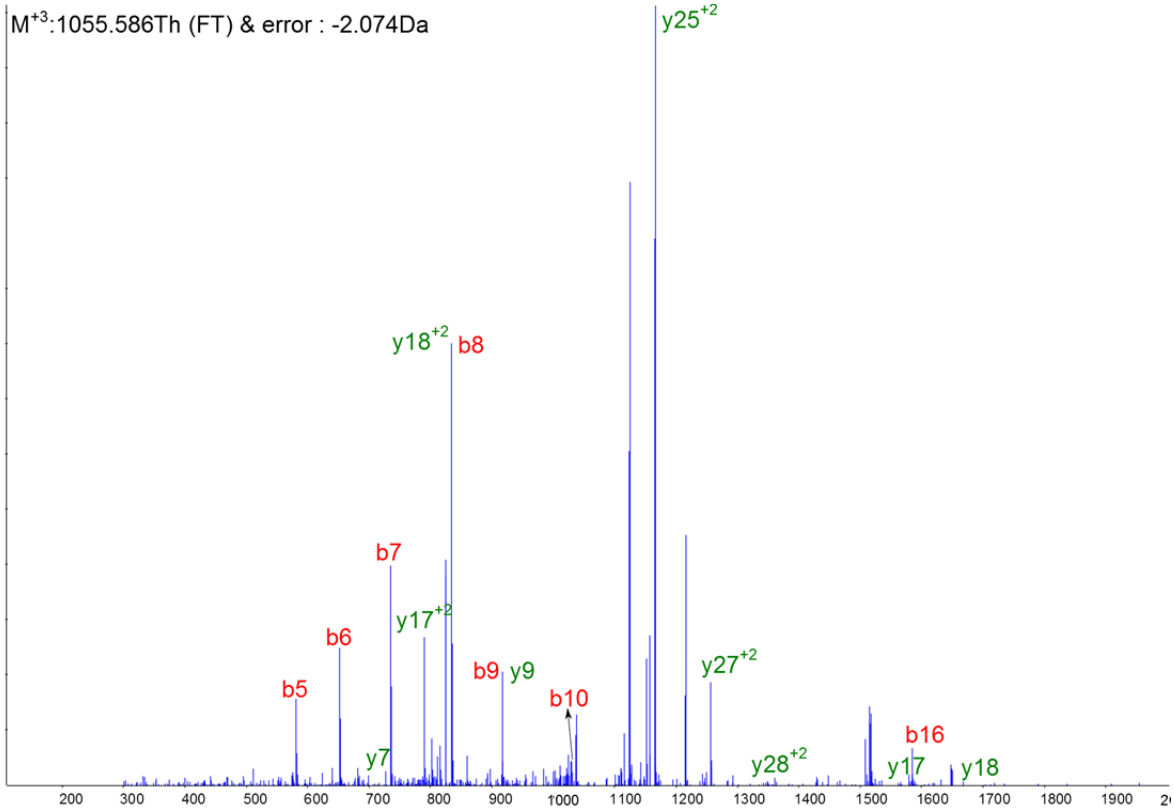
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
IT	527.105	1	b5	-0.164
IT	598.095	1	b6	-0.21
IT	681.177	1	b7	-0.169
IT	780.148	1	b8	-0.248
IT	863.35	1	b9	-0.088
IT	976.208	1	b10	-0.286
IT	1158.229	1	b12	-0.356
IT	1215.319	2	y26	0.201
IT	1461.248	1	b15	0.485
IT	1648.444	1	y18	-0.393
IT	1802.234	1	b19	0.327
IT	1951.684	1	y21	-0.304

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. CID spectrum of TTWACATVTLTVTVCSPTGTLCGSCSMGTRGCC (+2) with annotations from *S. roseosporus* NRRL 15998 (9 Dehyd).



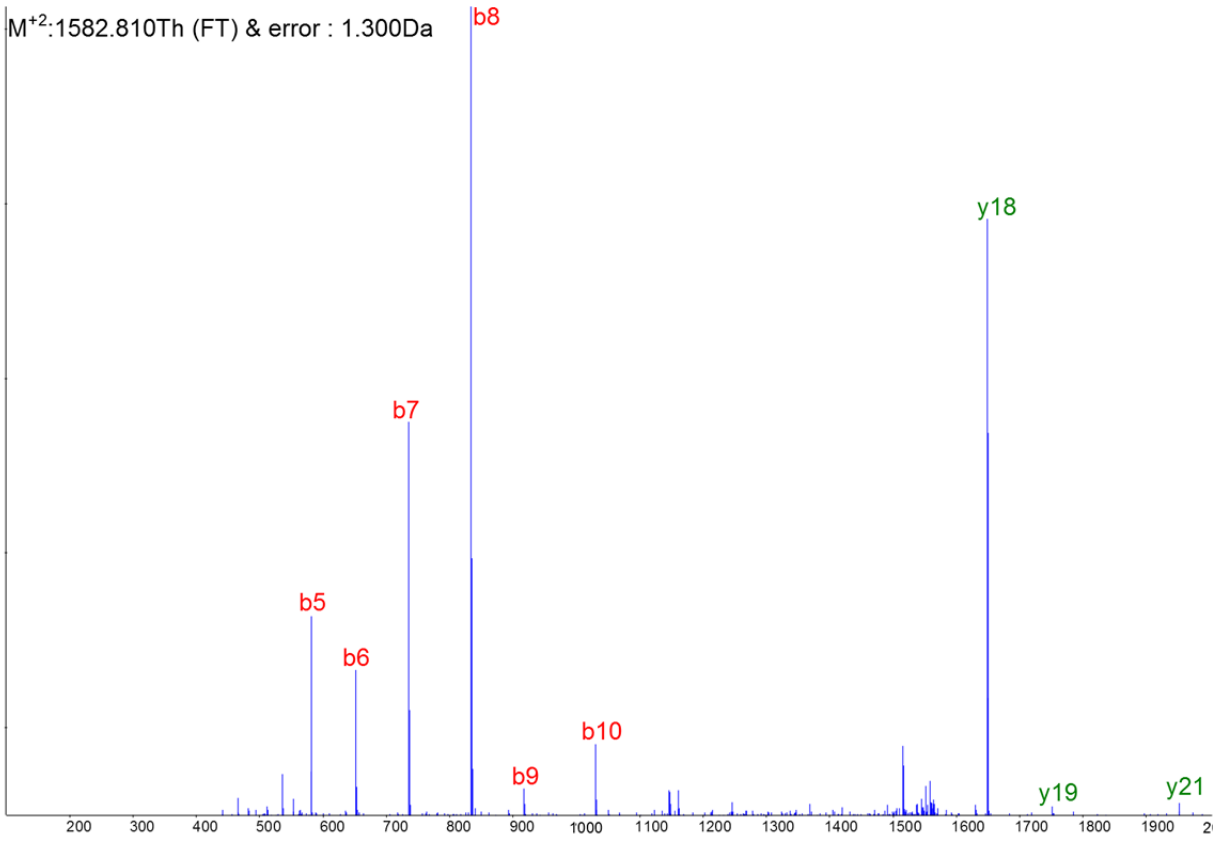
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
orbitrap	177.095	2	b3	-0.227
orbitrap	219.62	2	y4	0.09
orbitrap	527.27	1	b5	-0.062
orbitrap	681.347	1	b7	-0.065
orbitrap	709.368	1	y7	-0.478

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. HCD spectrum of TTWACATVTLTVTVCSPTGTLCGSCSMGTRGCC (+2) with annotations from *S. roseosporus* NRRL 15998 (8 Dehyd).



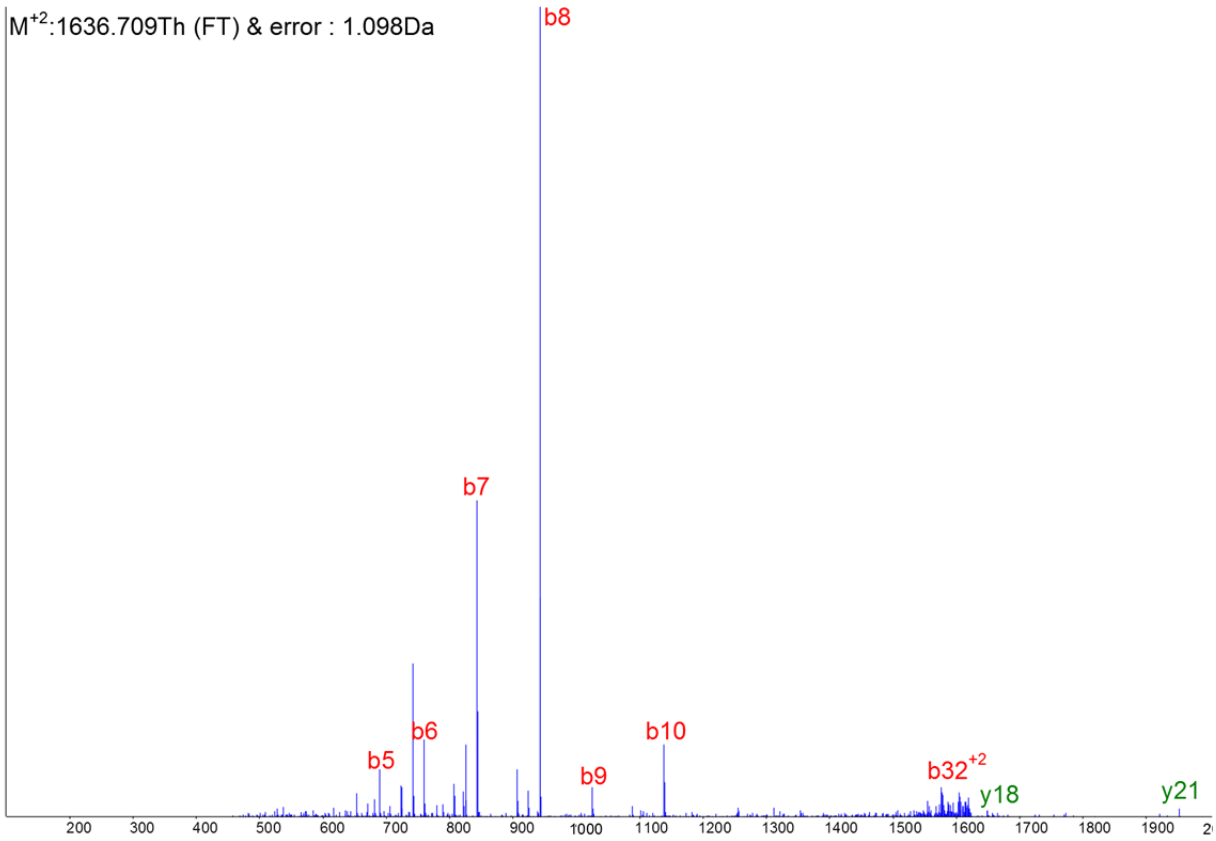
Detector	Observed m/z [Th]	Charge	Species	Difference (Th)
IT	580.898	1	b5	-0.398
IT	651.907	1	b6	-0.424
IT	727.369	1	y7	-0.007
IT	735.038	1	b7	-0.335
IT	790.327	2	y17	-0.077
IT	834.032	2	y18	0.105
IT	834.032	1	b8	-0.391
IT	916.992	1	b9	0.455
IT	916.922	1	y9	0.448
IT	1030.176	1	b10	-0.34
IT	1166.075	2	y25	-0.195
IT	1256.695	2	y27	0.056
IT	1292.101	2	y28	-0.055
IT	1579.657	1	y17	-0.145
IT	1584.349	1	b16	0.258
IT	1667.325	1	y18	0.479

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2.** CID spectrum of {TTWAC}$^{+54}$ATVTLTVTVCSPTGTLCGSCSMGTRGCC (+3) with annotations from *S. roseosporus* NRRL 15998 (9 Dehyd).



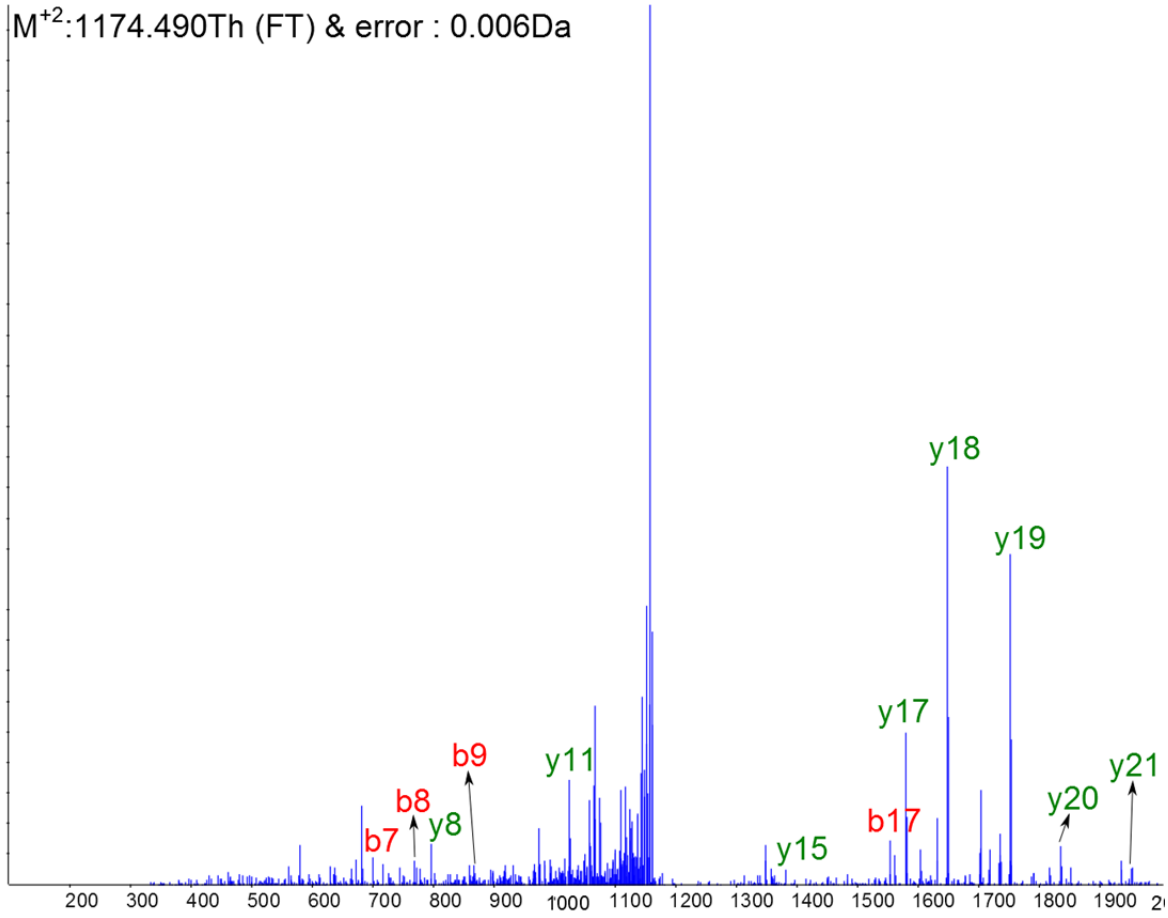
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
IT	581.098	1	b5	-0.199
IT	652.174	1	b6	-0.158
IT	735.137	1	b7	0.175
IT	834.161	1	b8	0.19
IT	917.063	1	b9	-0.401
IT	1030.219	1	b10	-0.302
IT	1648.384	1	y18	-0.453
IT	1751.52	2	y19	0.389
IT	1951.594	1	y21	-0.394

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. CID spectrum of {TTWAC}$^{+54}$ATVTLTVTVCSPTGTLCGSCSMGTRGCC (+2) with annotations from *S. roseosporus* NRRL 15998 (9 Dehyd).



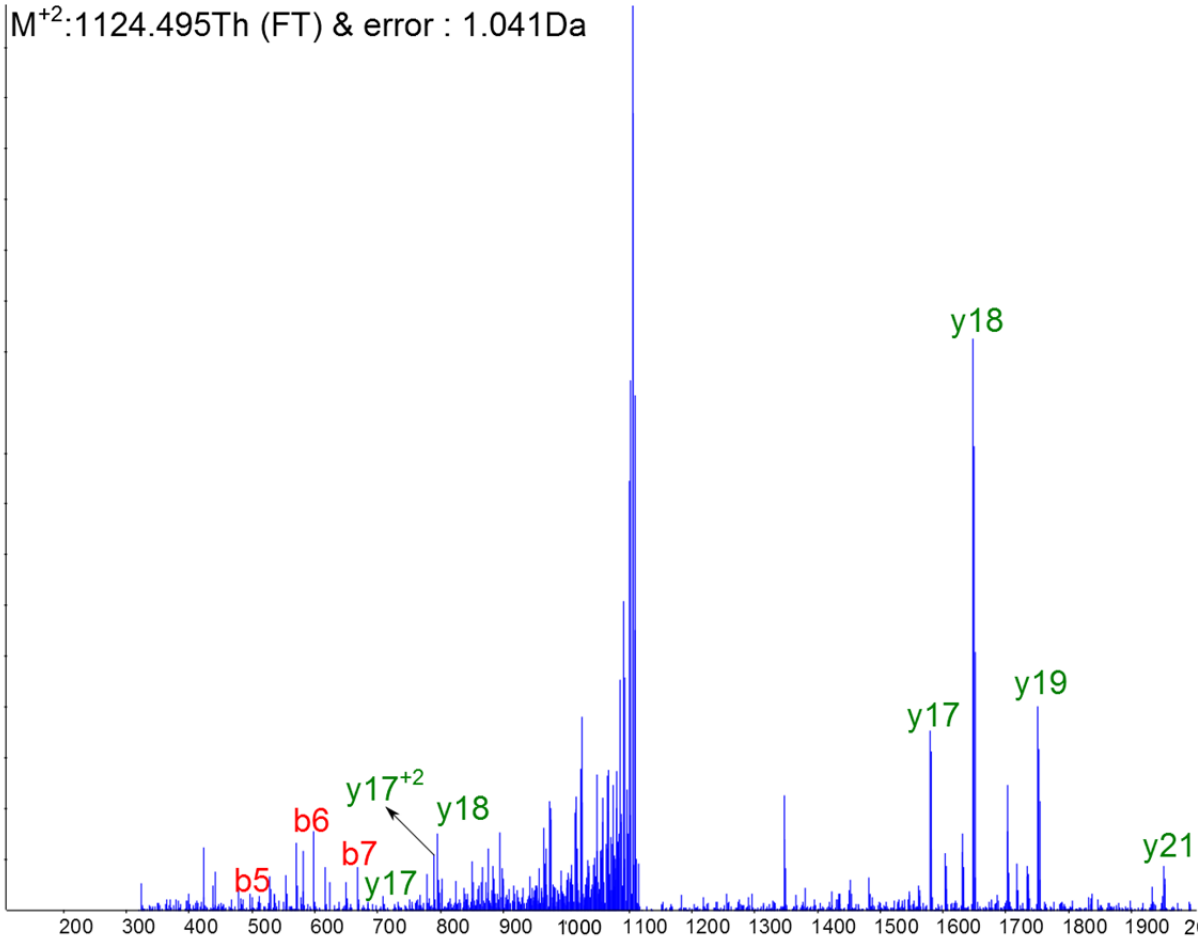
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
IT	689.011	1	b5	-0.339
IT	760.171	1	b6	-0.215
IT	843.197	1	b7	0.175
IT	942.167	1	b8	0.19
IT	1025.131	1	b9	-0.401
IT	1138.132	1	b10	-0.302
IT	1575.83	2	b32	0.036
IT	1648.544	1	y18	-0.292
IT	1951.637	1	y21	-0.351

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. CID spectrum of {TTWAC}$^{+162}$ATVTLTVTVCSPTGTLCGSCSMGTRGCC (+2) with annotations from *S. roseosporus* NRRL 15998 (9 Dehyd).



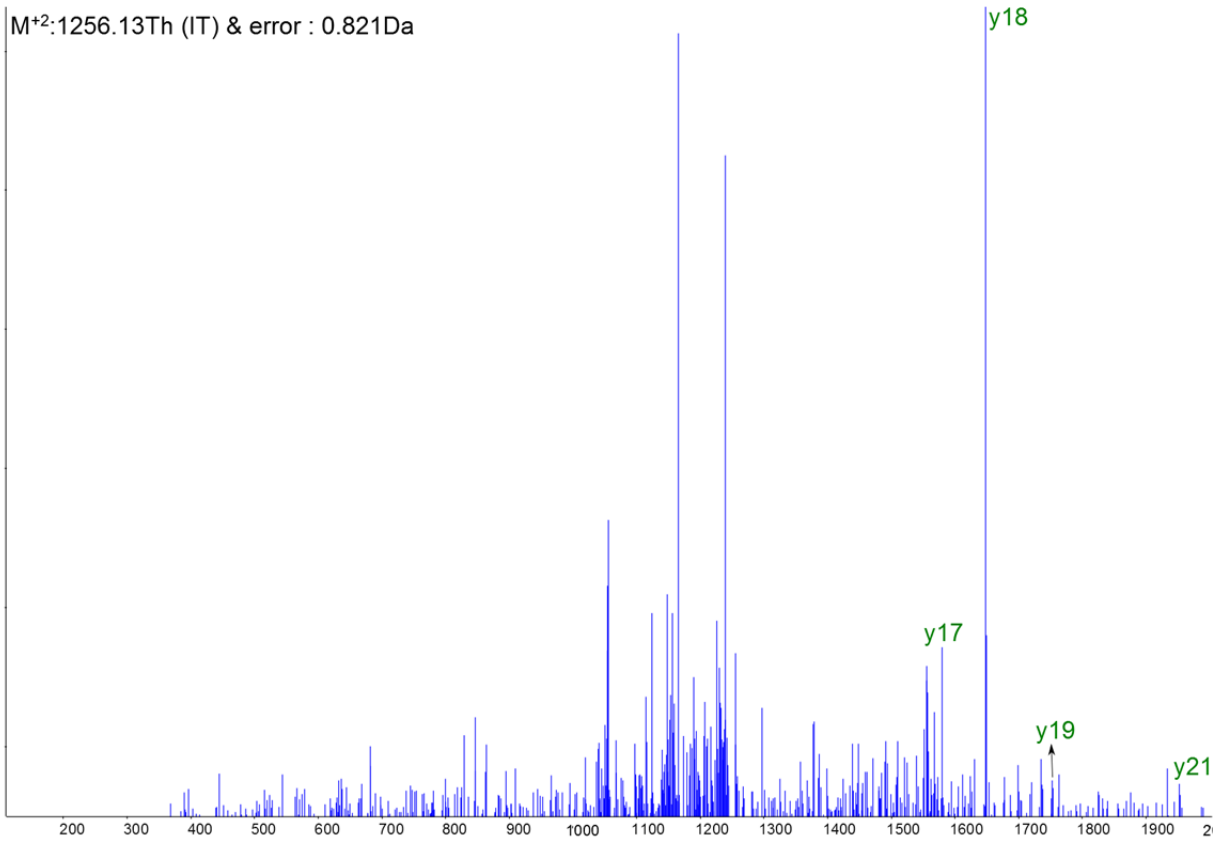
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
IT	700.237	1	b7	-0.119
IT	769.067	1	b8	-0.323
IT	796.24	1	y8	-0.17
IT	866.464	1	b9	0.024
IT	1025.064	1	y11	-0.461
IT	1381.377	1	y15	-0.326
IT	1449.373	1	b16	-0.357
IT	1552.471	1	b17	-0.312
IT	1579.387	1	y17	-0.415
IT	1648.346	1	y18	-0.491
IT	1751.417	1	y19	-0.471
IT	1850.485	1	y20	-0.453
IT	1951.676	1	y21	-0.312

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. CID spectrum of TLTVTVCSPTGTLCGSCSMGTRGCC (+2) with annotations from *S. roseosporus* NRRL 15998 (5 Dehyd).



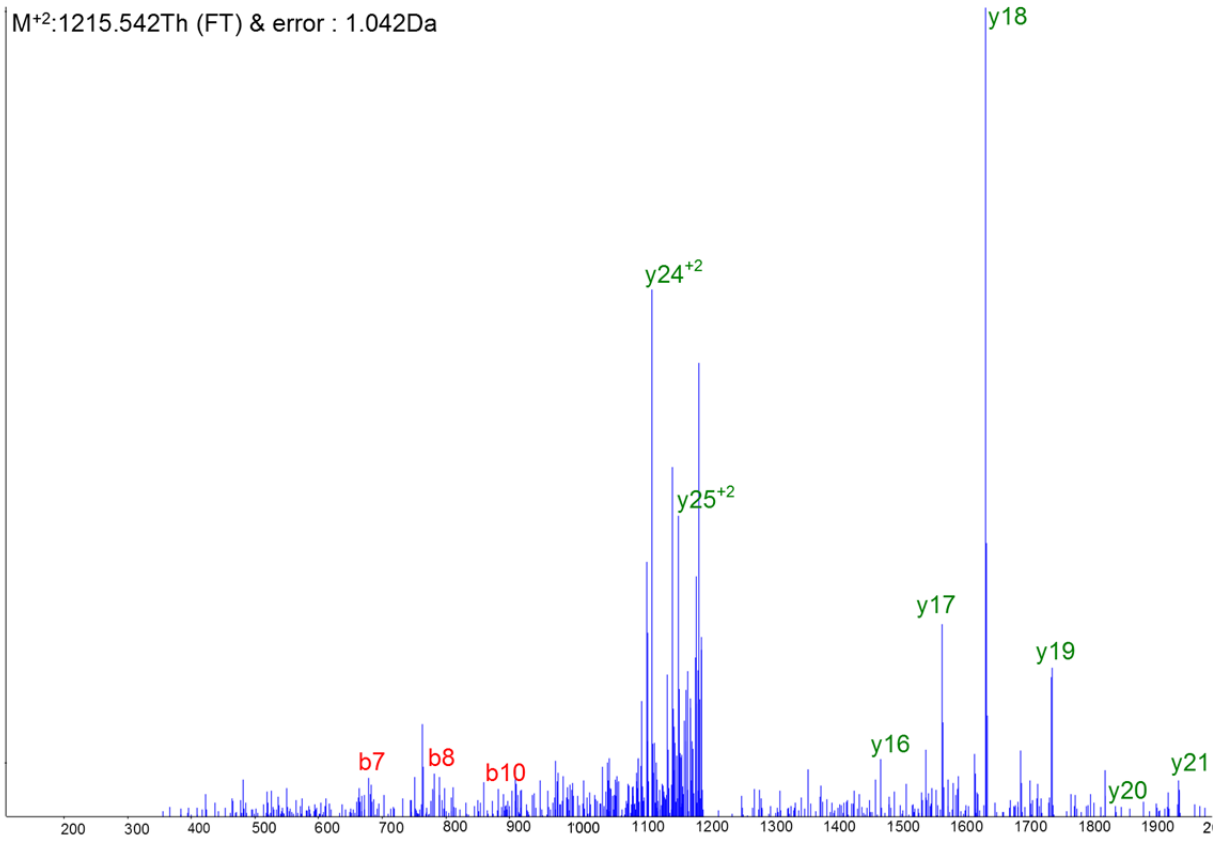
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
IT	496.172	1	b5	-0.082
IT	599.184	1	b6	-0.121
IT	668.185	1	b7	-0.155
IT	709.157	1	y7	-0.21
IT	790.464	2	y17	0.059
IT	796.008	1	y8	-0.402
IT	1579.337	1	y17	-0.466
IT	1648.414	1	y18	-0.423
IT	1751.475	1	y19	-0.413
IT	1951.494	1	y21	-0.495

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. CID spectrum of LTVTVCSPTGTLCGSCSMGTRGCC (+2) with annotations from *S. roseosporus* NRRL 15998 (5 Dehyd).



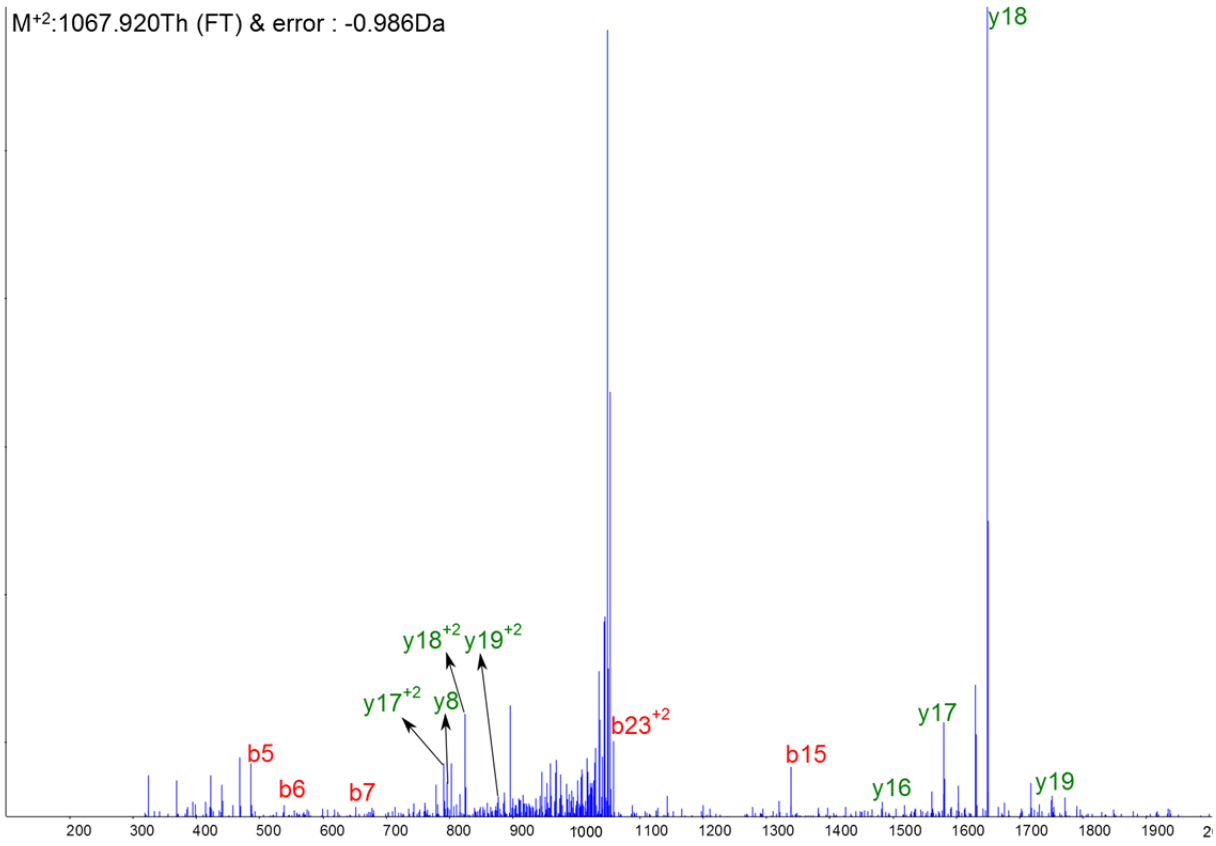
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
IT	1579.456	1	y17	-0.347
IT	1648.476	1	y18	0.529
IT	1751.671	1	y19	0.97
IT	1951.637	1	y21	0.957

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. CID spectrum of TVTLTVTVCSPTGTLCGSCSMGTRGCC (+2) with annotations from *S. roseosporus* NRRL 15998 (7 Dehyd).



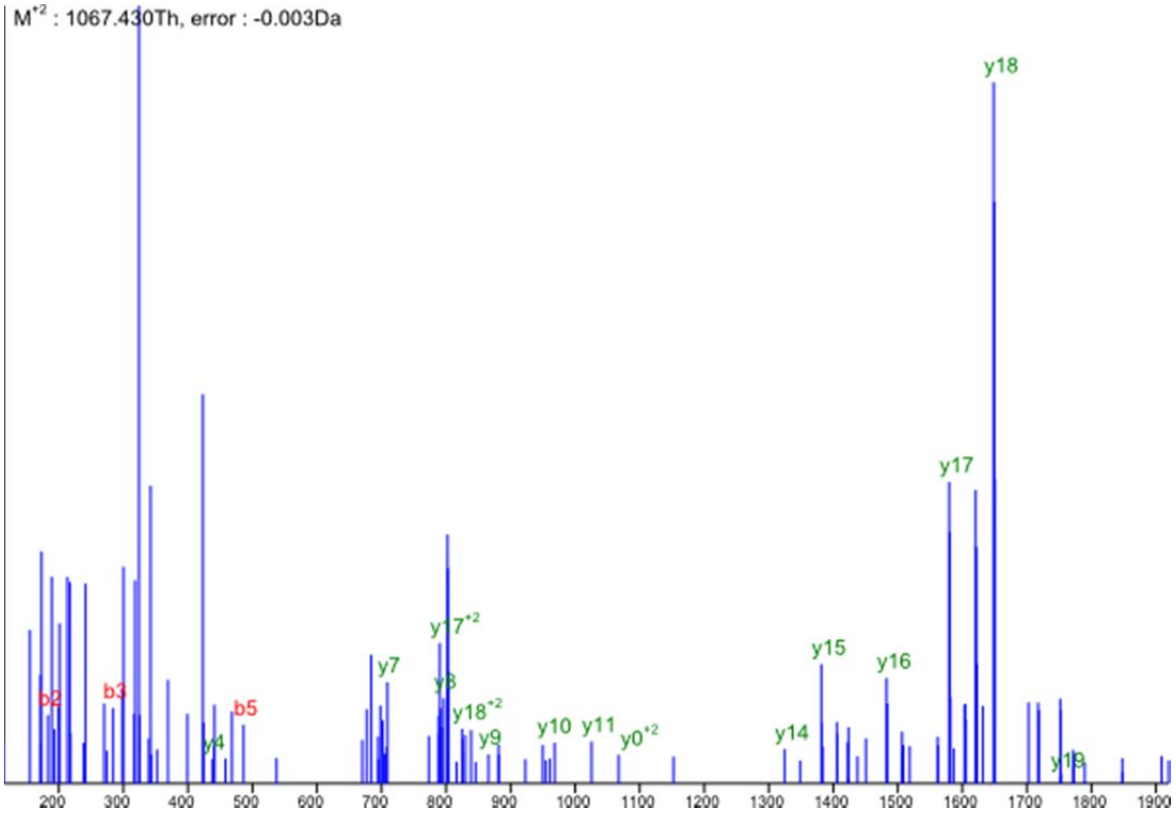
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
IT	678.315	1	b7	-0.03
IT	781.522	1	b8	0.942
IT	947.487	1	b10	0.007
IT	1124.019	2	y24	-0.053
IT	1165.607	2	y25	0.014
IT	1482.566	1	y16	-0.188
IT	1579.54	1	y17	0.801
IT	1648.39	1	y18	0.59
IT	1751.512	1	y19	0.598
IT	1852.196	1	y20	0.263
IT	1950.614	1	y21	0.732

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. CID spectrum of VTLTVTVCSPTGTLCGSCSMGTRGCC (+2) with annotations from *S. roseosporus* NRRL 15998 (6 Dehyd).



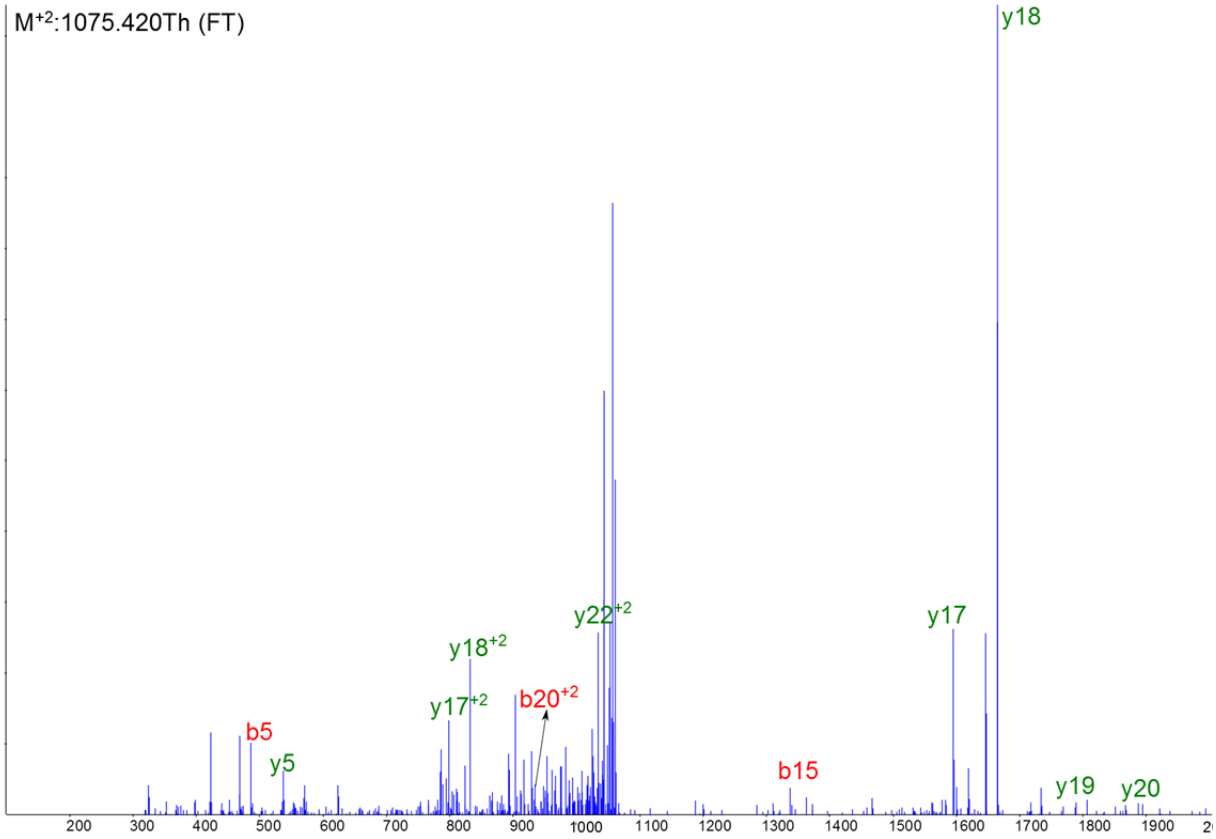
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
IT	486.162	1	b5	-0.087
IT	554.824	1	b6	-0.449
IT	652.216	1	b7	-0.116
IT	790.357	2	y17	-0.047
IT	796.095	1	y8	-0.316
IT	824.679	2	y18	0.412
IT	876.46	2	y19	0.012
IT	1058.602	2	b23	0.066
IT	1338.245	1	b15	0.654
IT	1482.445	1	y16	-0.309
IT	1579.457	1	y17	0.39
IT	1648.428	1	y18	0.62
IT	1751.369	1	y19	-0.519

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. CID spectrum of TVTVCSPTGTLCGSCSMGTRGCC (+2) with annotations from *S. roseosporus* NRRL 15998 (5 Dehyd).



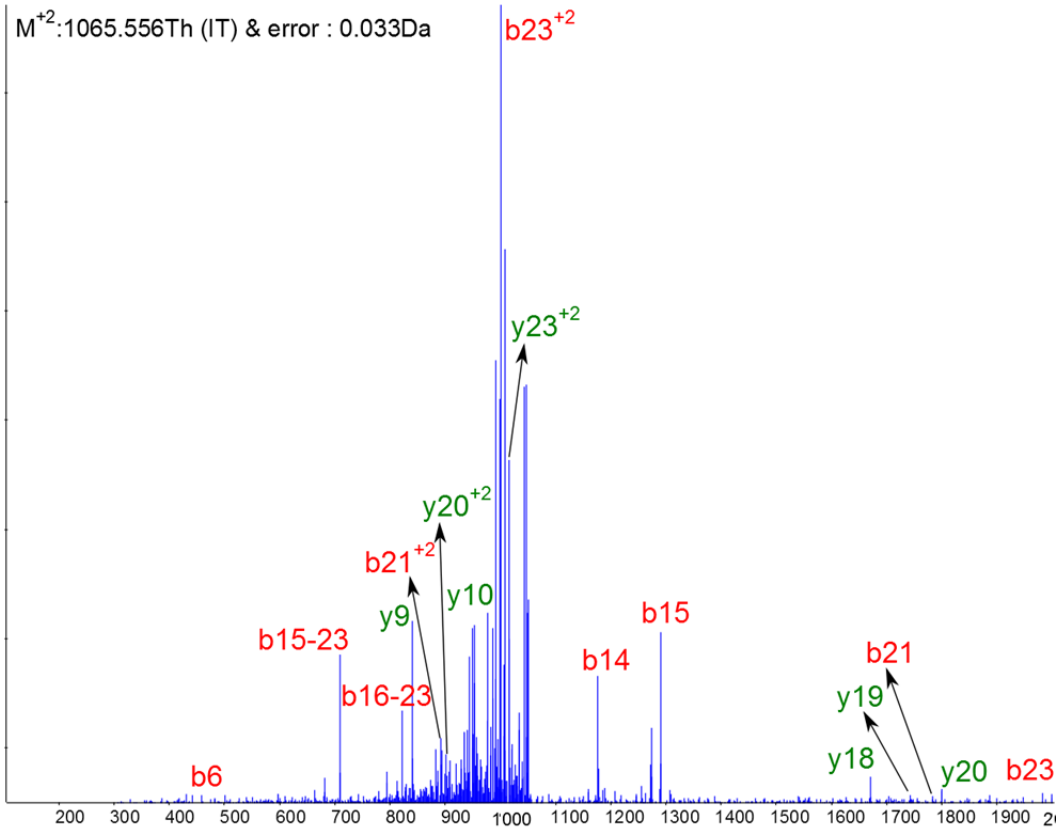
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
orbitrap	183.098	1	b2	0.013
orbitrap	284.149	1	b3	0.013
orbitrap	438.232	1	y4	-0.051
orbitrap	486.25	1	b5	-0.007
orbitrap	709.368	1	y7	-0.249
orbitrap	790.405	2	y17	-0.106
orbitrap	796.411	1	y8	-0.12
orbitrap	824.922	2	y18	-0.112
orbitrap	865.446	1	y9	0.356
orbitrap	968.497	1	y10	-0.201
orbitrap	1025.53	1	y11	-0.175
orbitrap	1067.54	2	y0	-0.134
orbitrap	1324.68	1	y14	-0.188
orbitrap	1381.7	1	y15	-0.223
orbitrap	1482.75	1	y16	-0.213
orbitrap	1579.8	1	y17	-0.224
orbitrap	1648.84	1	y18	-0.243
orbitrap	1751.89	1	y19	0.065

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. HCD spectrum of TVTVCSPTGTLCGSCSMGTRGCC (+2) with annotations from *S. roseosporus* NRRL 15998 (5 Dehyd).



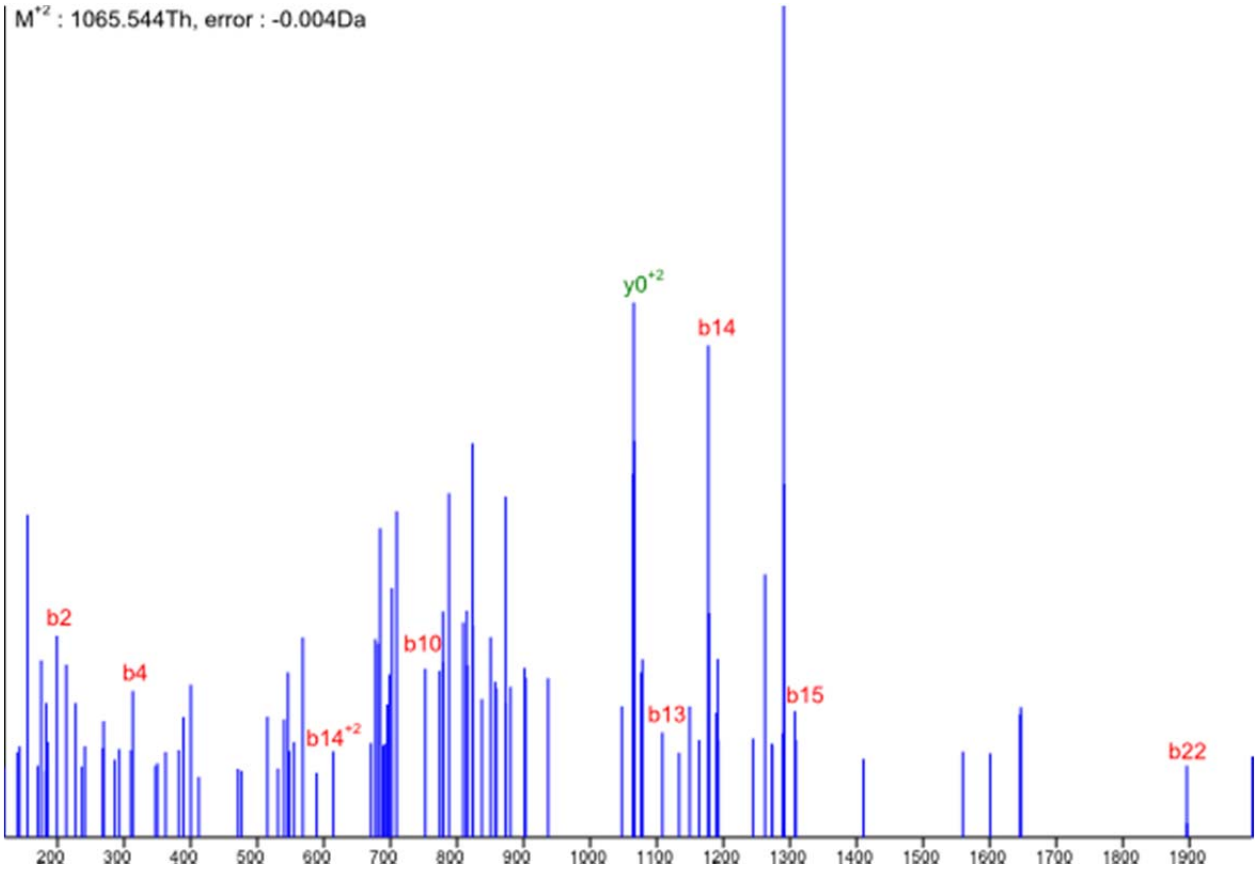
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
IT	486.15	1	b5	-0.068
IT	537.037	1	y5	0.244
IT	798.489	2	y17	0.08
IT	832.979	2	y18	0.052
IT	934.944	2	b20	-0.03
IT	1034.478	2	y22	0.451
IT	1338.173	1	b15	-0.502
IT	1595.431	1	y17	0.675
IT	1664.38	1	y18	0.513
IT	1768.552	1	y19	0.513
IT	1867.058	1	y20	0.112

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. CID spectrum of TVTVCSPTGTLCGSCSMGTRGCC (+2) with annotations from *S. roseosporus* NRRL 15998 (4 Dehyd).



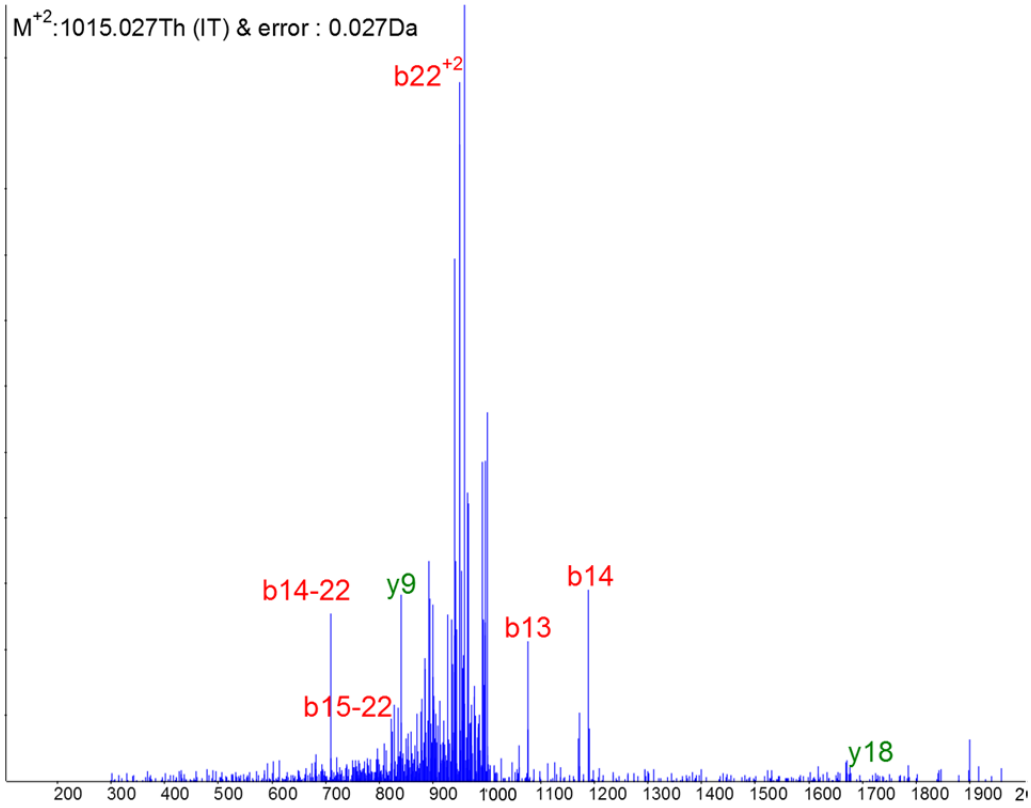
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
IT	459.192	1	b6	-0.044
IT	709.244	1	b15-23	-0.107
IT	822.284	1	b16-23	-0.14
IT	840.343	1	y9	-0.089
IT	892.069	2	b21	0.116
IT	900.64	2	y20	0.18
IT	953.508	1	y10	0.018
IT	1000.475	2	b23	-0.213
IT	1015.008	2	y23	-0.009
IT	1177.239	1	b14	-0.355
IT	1290.422	1	b15	-0.229
IT	1671.623	1	y18	-0.225
IT	1742.797	1	y19	-0.087
IT	1783.006	1	b21	0.108
IT	1799.717	1	y20	-0.196
IT	1999.605	1	b23	-0.143

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. CID spectrum of TDGGGASTVSLLSCISAASVLLCL(+2) with annotations from *S. viridochromogenes* DSM 40736 (6 Dehyd).



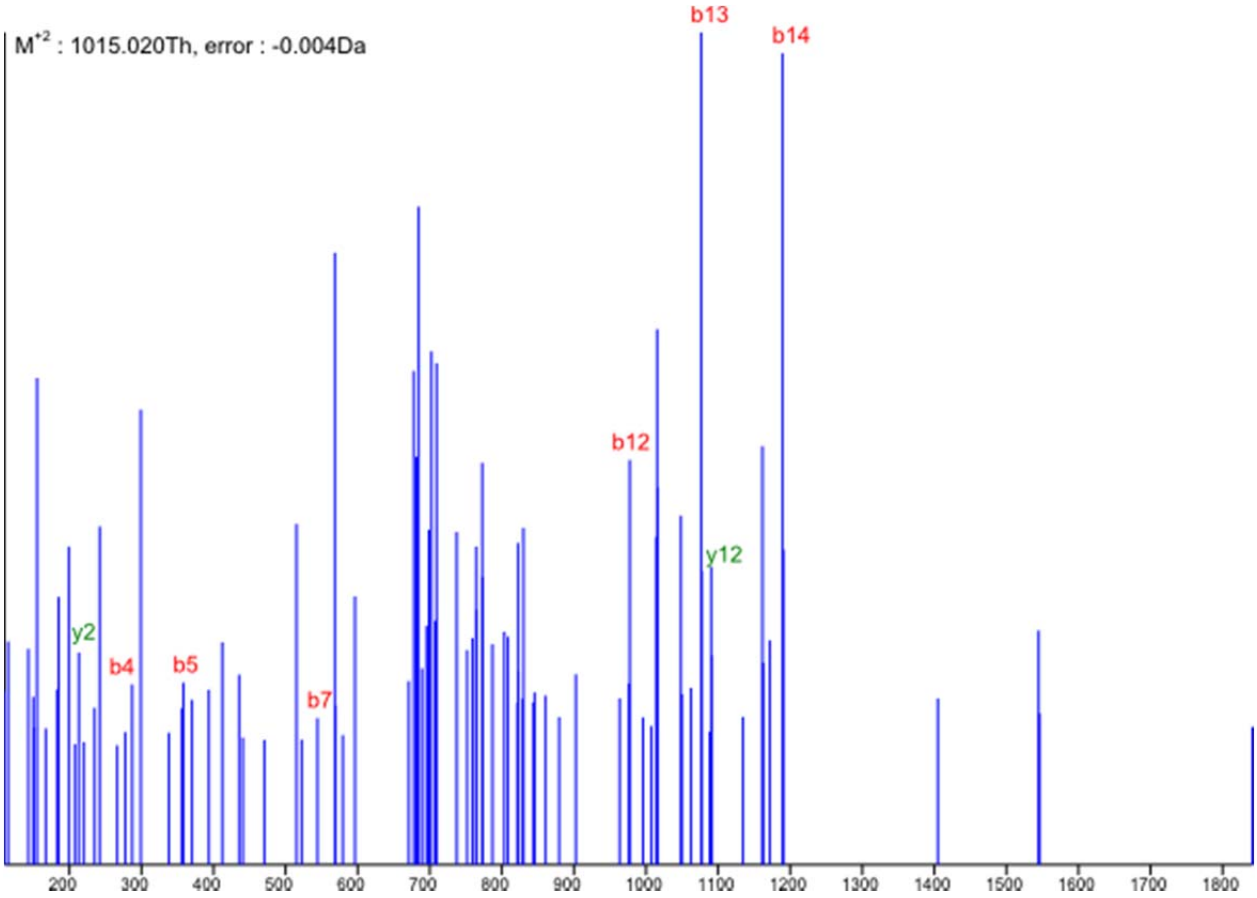
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
orbitrap	199.106	1	b2	0.074
orbitrap	313.163	1	b4	-0.071
orbitrap	589.301	2	b14	-0.047
orbitrap	779.396	1	b10	-0.044
orbitrap	1065.54	2	y0	-0.066
orbitrap	1108.56	1	b13	-0.028
orbitrap	1177.6	1	b14	-0.078
orbitrap	1290.65	1	b15	0.238
orbitrap	1895.95	1	b22	0.454

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. HCD spectrum of TDGGGASTVSLLSCISAASVLLCL(+2) with annotations from *S. viridochromogenes* DSM 40736 (6 Dehyd).



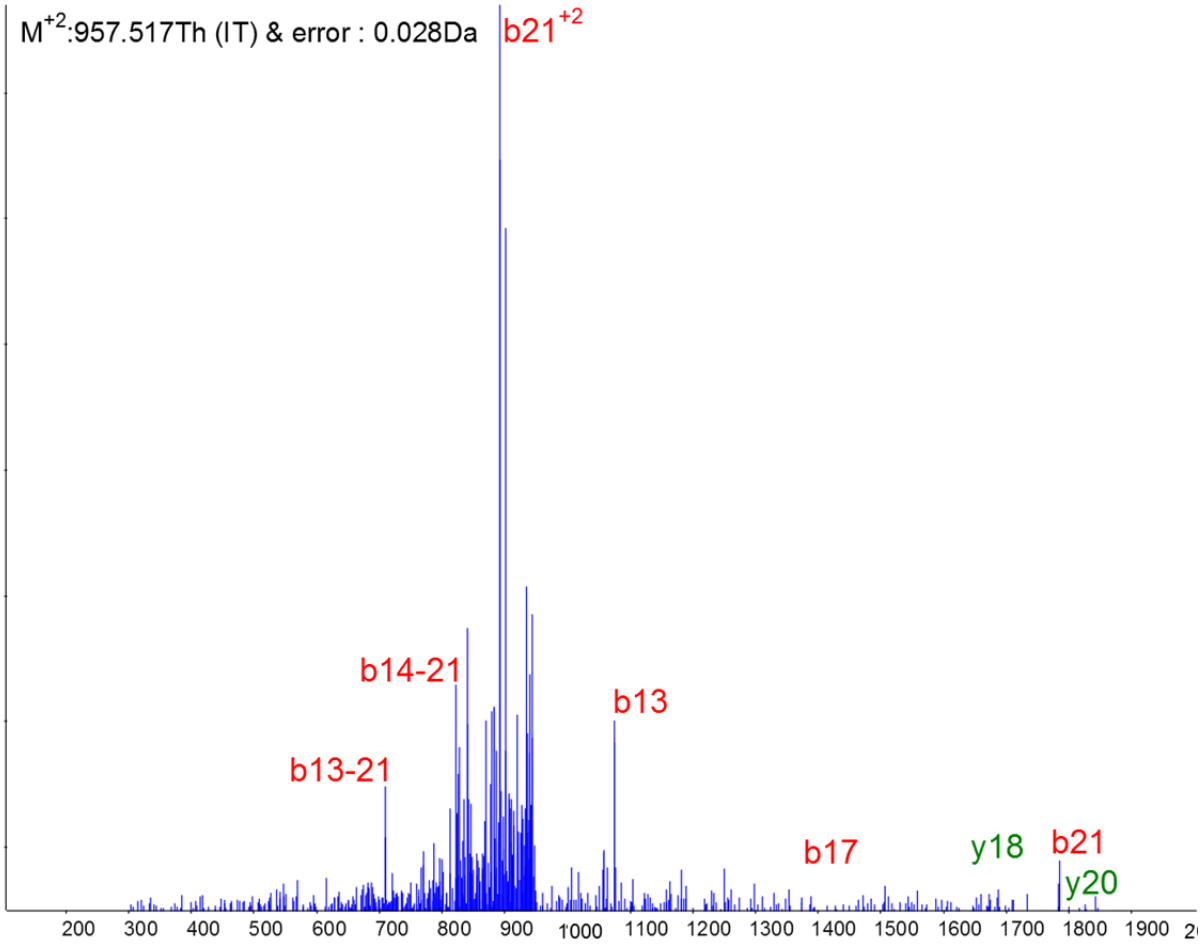
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
IT	459.192	1	b6	-0.044
IT	709.244	1	b15-23	-0.107
IT	822.284	1	b16-23	-0.14
IT	840.343	1	y9	-0.089
IT	892.069	2	b21	0.116
IT	900.64	2	y20	0.18
IT	953.508	1	y10	0.018
IT	1000.475	2	b23	-0.213
IT	1015.008	2	y23	-0.009
IT	1177.239	1	b14	-0.355
IT	1290.422	1	b15	-0.229
IT	1671.623	1	y18	-0.225
IT	1742.797	1	y19	-0.087
IT	1783.006	1	b21	0.108
IT	1799.717	1	y20	-0.196
IT	1999.605	1	b23	-0.143

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. CID spectrum of DGGGASTVSLLSCISAASVLLCL(+2) with annotations from *S. viridochromogenes* DSM 40736 (6 Dehyd).



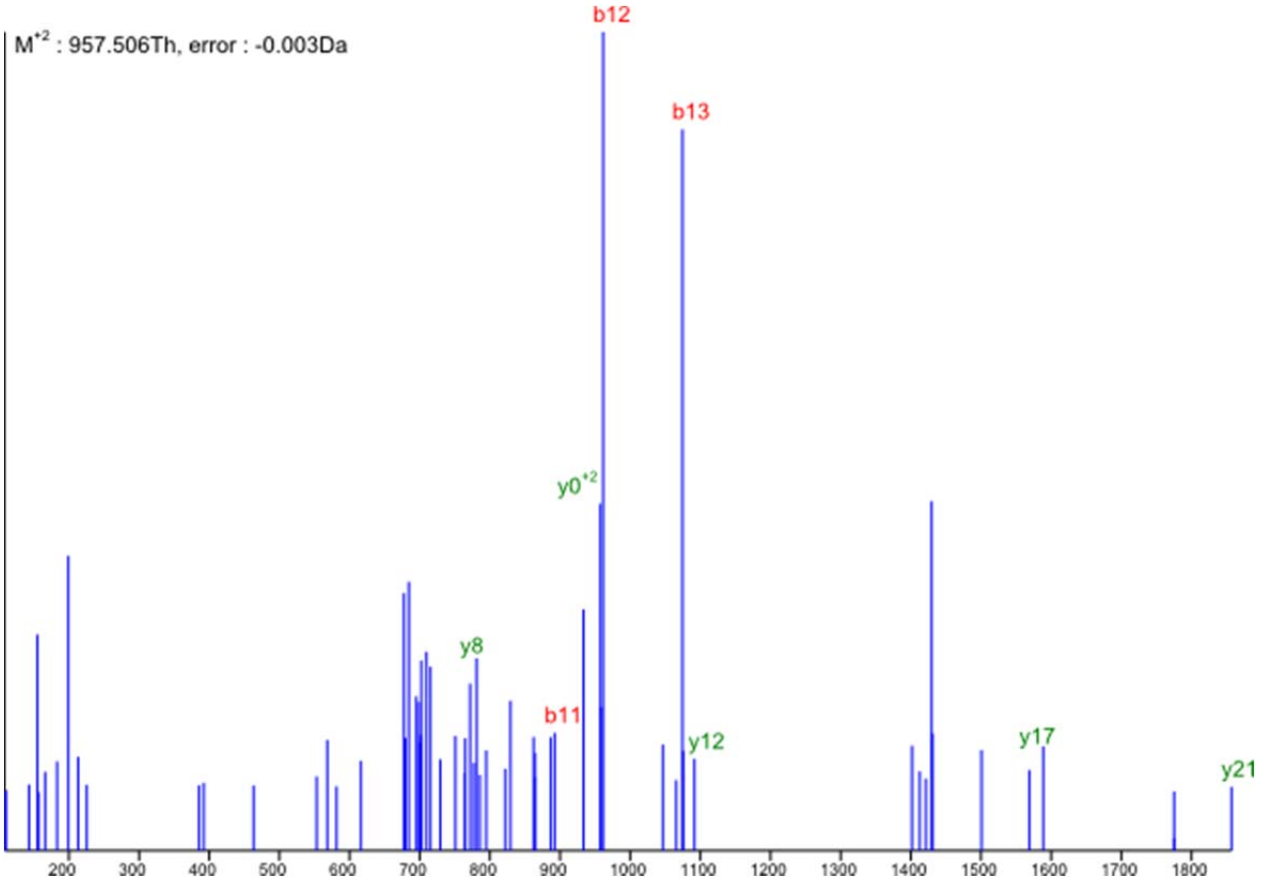
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
orbitrap	235.131	1	y2	-0.294693
orbitrap	287.15	1	b4	-0.0471497
orbitrap	358.186	1	b5	-0.0503235
orbitrap	544.279	1	b7	-0.0534668
orbitrap	1007.51	1	b12	-0.0648193
orbitrap	1076.55	1	b13	-0.0698242
orbitrap	1091.56	1	y12	-0.0540771
orbitrap	1189.6	1	b14	-0.0354004

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. HCD spectrum of DGGGASTVSLLSCISAASVLLCL(+2) with annotations from *S. viridochromogenes* DSM 40736 (6 Dehyd).



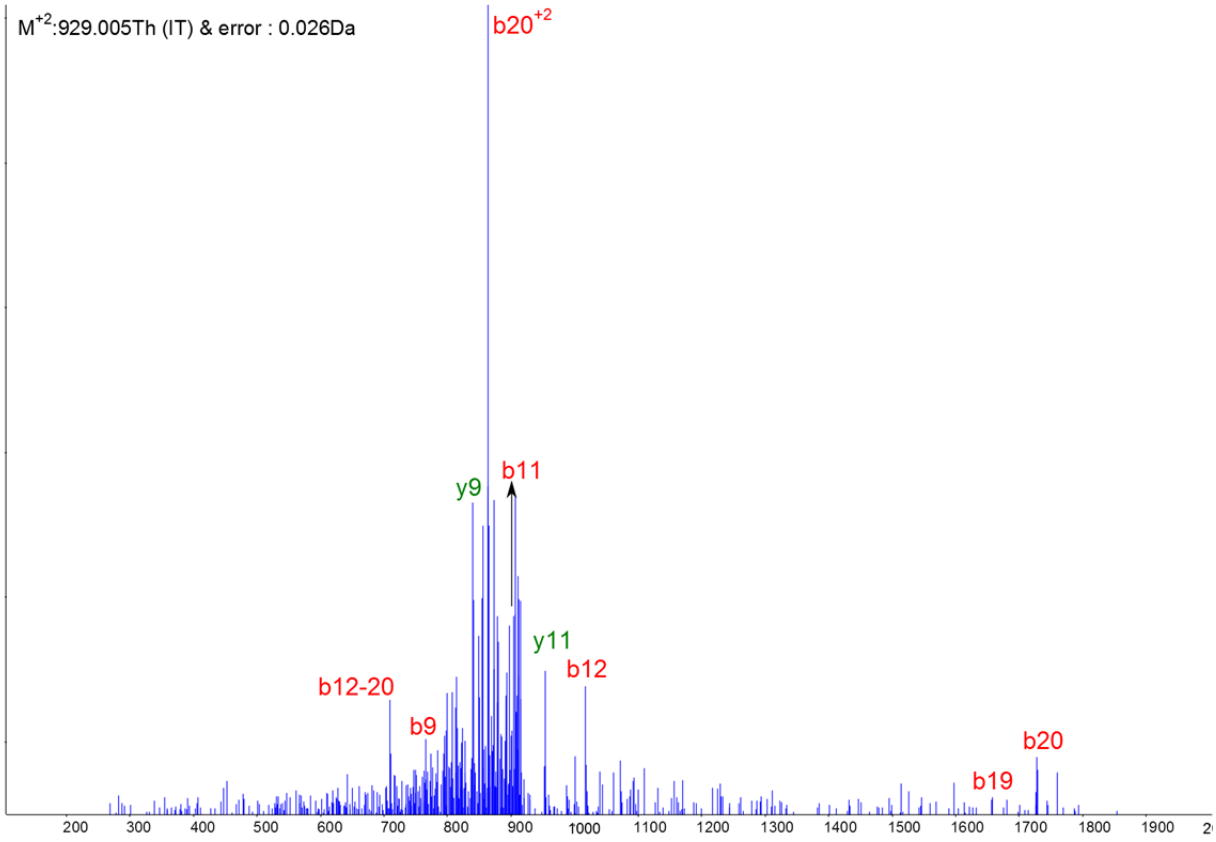
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
IT	709.369	1	b13-21	0.018
IT	822.133	1	b14-22	-0.151
IT	892.1	2	b21	0.147
IT	1074.403	1	b13	-0.141
IT	1388.306	1	b17	-0.394
IT	1671.83	1	y18	-0.018
IT	1782.81	1	b21	-0.08
IT	1799.85	1	y20	-0.063

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. CID spectrum of GGGASTVSLLSCISAASVLLCL(+2) with annotations from *S. viridochromogenes* DSM 40736 (6 Dehyd).



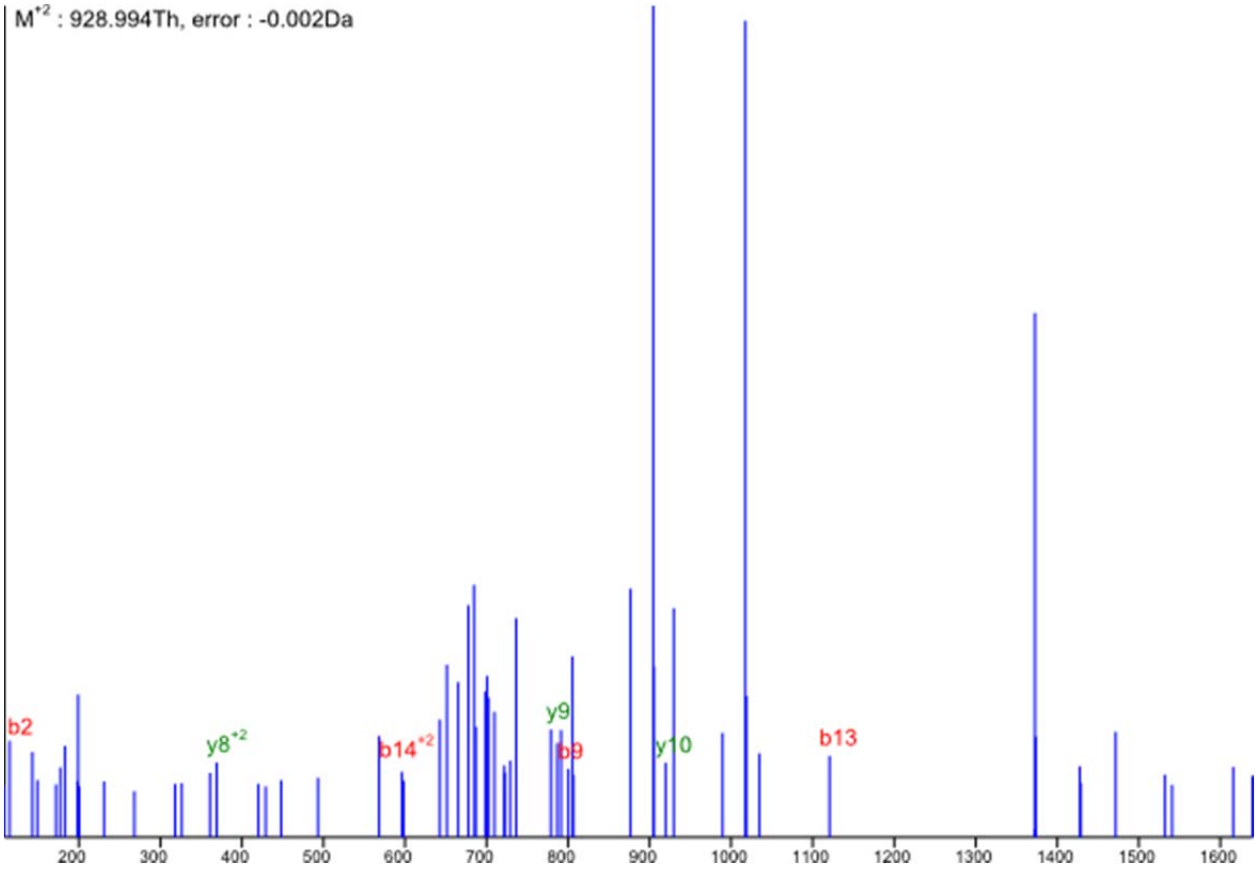
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
orbitrap	771.399	1	y8	0.487
orbitrap	892.453	1	b11	-0.017
orbitrap	957.489	2	y0	-0.024
orbitrap	961.488	1	b12	-0.034
orbitrap	1074.54	1	b13	-0.001
orbitrap	1091.56	1	y12	0.011
orbitrap	1568.8	1	y17	0.095
orbitrap	1856.94	1	y21	0.126

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. HCD spectrum of GGGASTVSLLSCISAASVLLCL(+2) with annotations from *S. viridochromogenes* DSM 40736 (6 Dehyd).



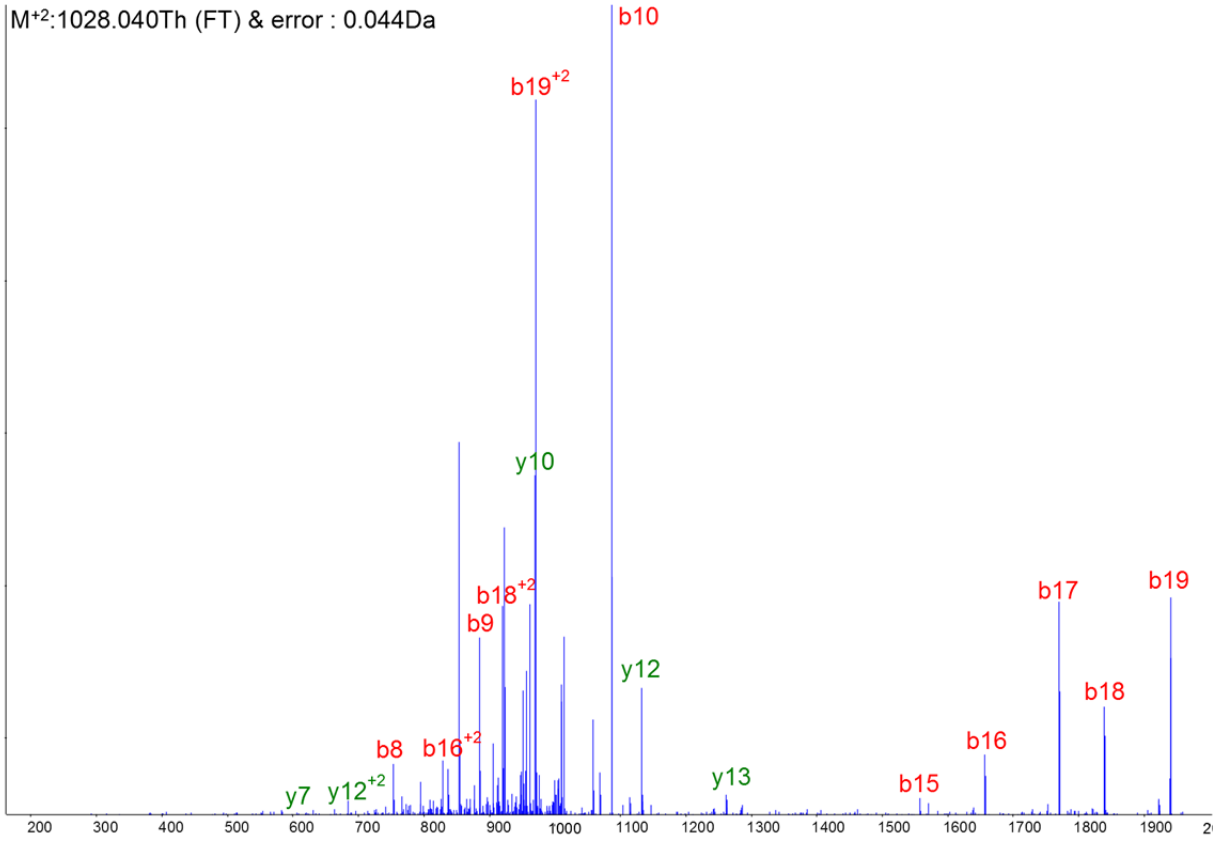
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
IT	709.172	1	b12-20	-0.179
IT	766.329	1	b9	-0.059
IT	840.319	1	y9	-0.113
IT	863.585	2	b20	0.147
IT	904.356	1	b11	-0.102
IT	953.609	1	y11	0.119
IT	1017.424	1	b12	-0.091
IT	1656.238	1	b19	0.424
IT	1725.946	1	b20	0.076

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. CID spectrum of GGASTVSLLSCISAASVLLCL(+2) with annotations from *S. viridochromogenes* DSM 40736 (6 Dehyd).



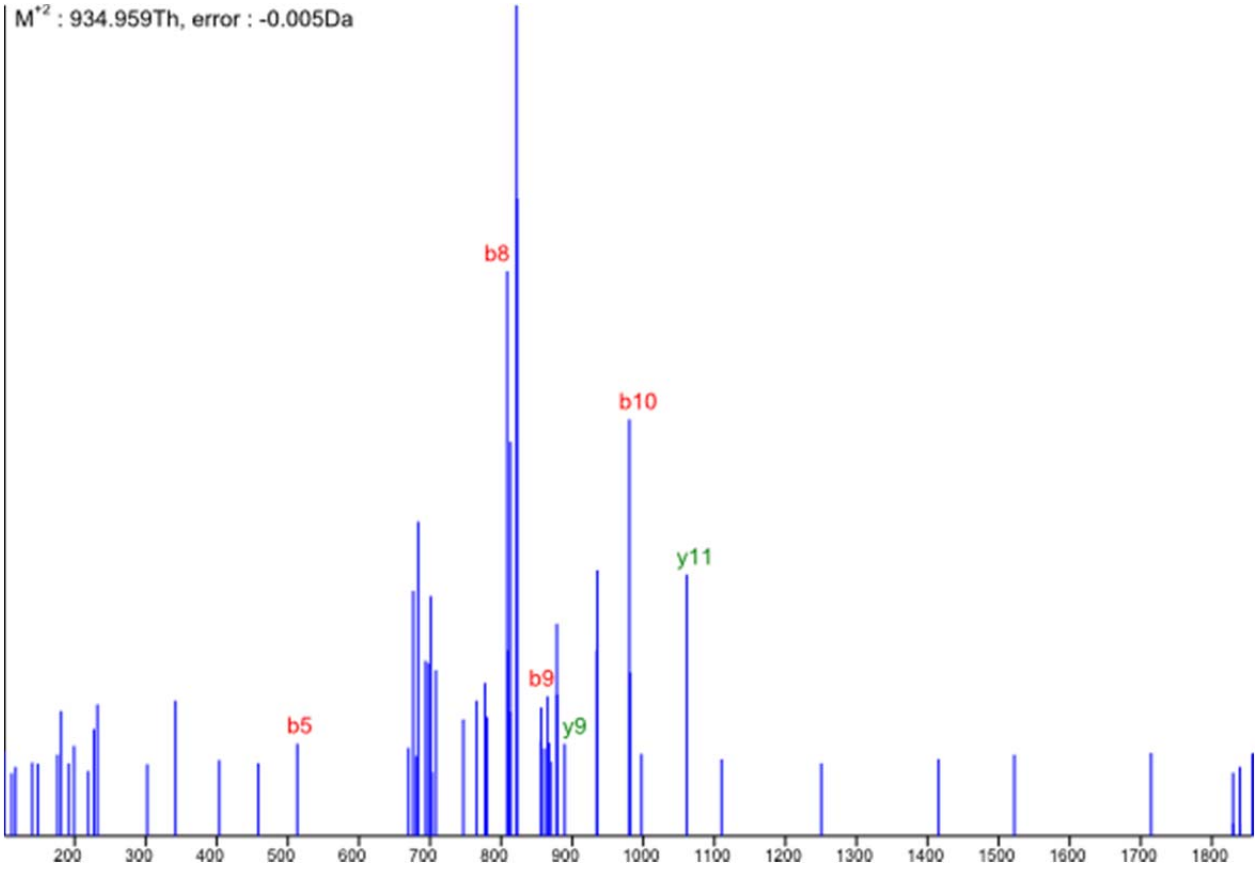
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
orbitrap	115.064	1	b2	0.022
orbitrap	369.195	2	y8	-0.073
orbitrap	596.305	2	b14	-0.003
orbitrap	800.407	1	b9	0.003
orbitrap	806.416	1	y9	-0.061
orbitrap	919.473	1	y10	0.152
orbitrap	1120.57	1	b13	0.193

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. HCD spectrum of GGASTVSLLSCISAASVLLCL(+2) with annotations from *S. viridochromogenes DSM 40736* (6 Dehyd).



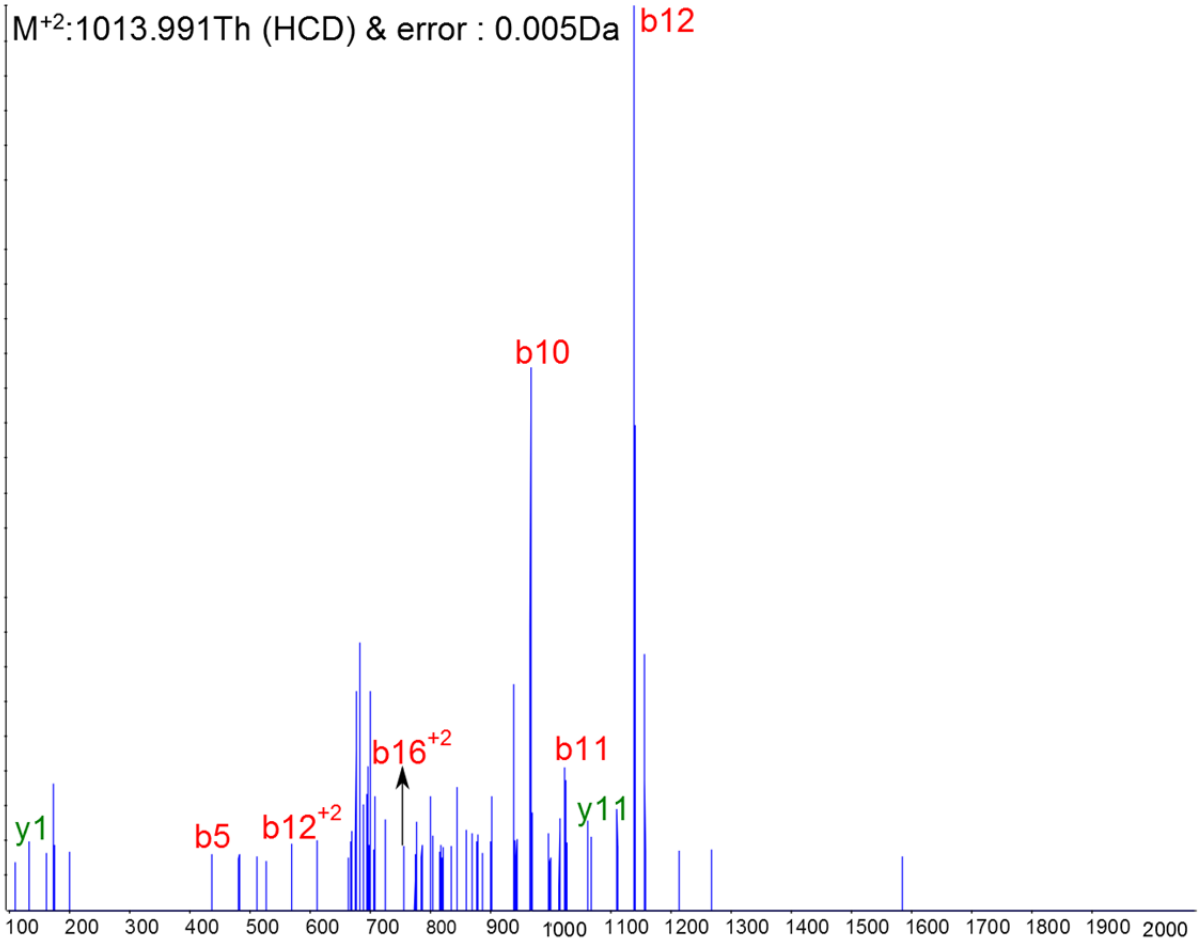
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
IT	631.461	2	y12	0.136
IT	684.31	1	y7	-0.045
IT	794.46	1	b8	0.056
IT	829.028	2	b16	0.107
IT	885.469	1	b9	0.02
IT	920.124	2	b18	0.157
IT	969.241	1	y10	-0.256
IT	970.546	2	b19	0.054
IT	1086.333	1	b10	-0.216
IT	1132.373	1	y12	-0.205
IT	1261.433	1	y13	-0.21
IT	1557.407	1	b15	-0.377
IT	1656.709	1	b16	-0.126
IT	1769.695	1	b17	-0.196
IT	1838.69	1	b18	-0.235
IT	1939.897	1	b19	-0.079

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. CID spectrum of GSQVSLLVCEYSSLSVVLCTP (+2) with annotations from *S. griseus* IFO 13350 (4 Dehyd).



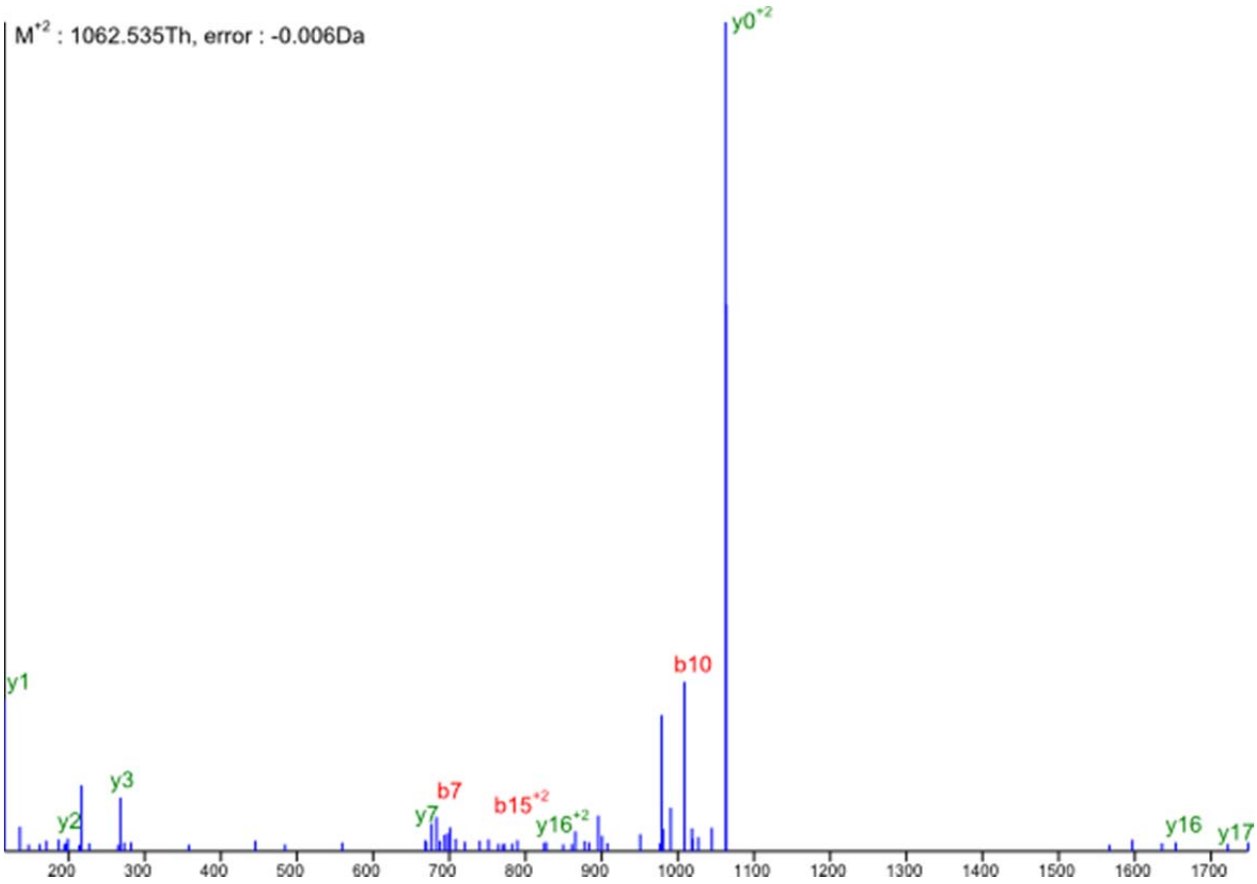
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
orbitrap	513.263	1	b5	-0.001
orbitrap	808.411	1	b8	0.045
orbitrap	865.439	1	b9	0.055
orbitrap	889.458	1	y9	-0.055
orbitrap	980.497	1	b10	0.008
orbitrap	1061.54	1	y11	-0.09

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. HCD spectrum of SRASLLLCGDSSLSITTCN (+2) with annotations from *S. lividans* TK24 (4 Dehyd).



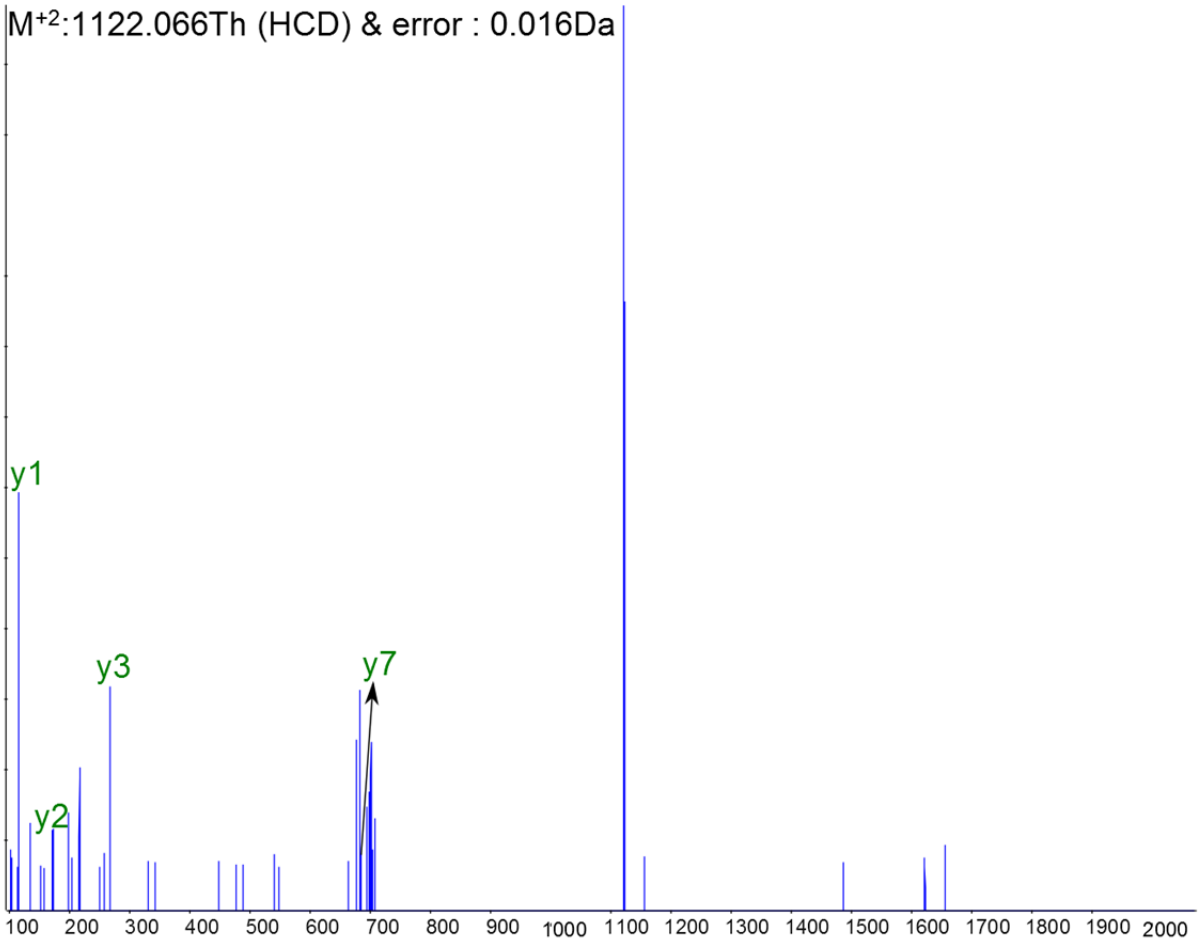
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
orbitrap	133.06	1	y1	-0.018
orbitrap	437.278	1	b5	0.05
orbitrap	569.731	2	b12	-0.051
orbitrap	755.803	2	b16	0.065
orbitrap	966.538	1	b10	0.033
orbitrap	1023.536	1	b11	0.022
orbitrap	1061.484	1	y11	-0.068
orbitrap	1138.575	1	b12	0.008

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. HCD spectrum of TGSRASLLLCGDSSLSITTCN (+2) with annotations from *S. lividans* TK24 (4 Dehyd).



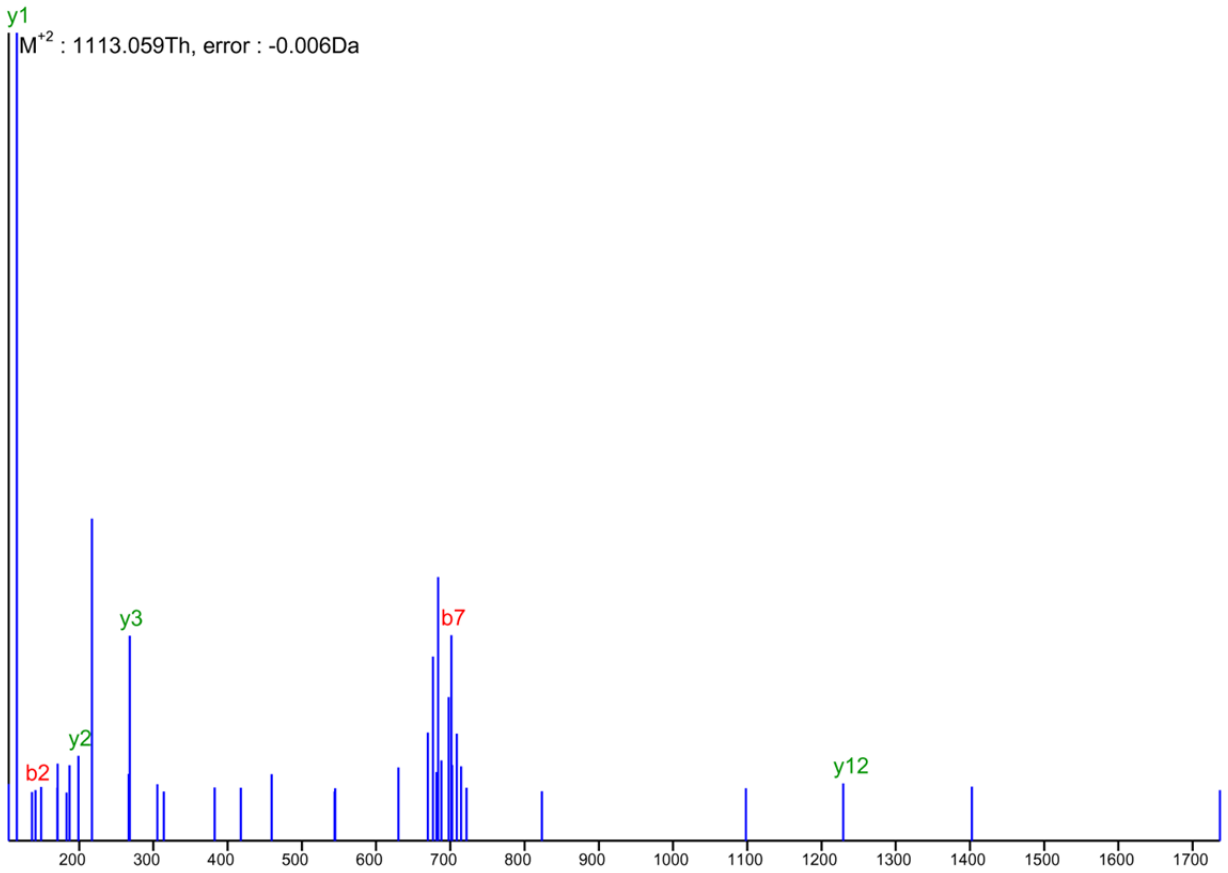
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
orbitrap	116.071	1	y1	0.009
orbitrap	199.113	1	y2	-0.003
orbitrap	268.147	1	y3	-0.017
orbitrap	668.347	1	y7	0.07
orbitrap	697.355	1	b7	-0.028
orbitrap	772.393	2	b15	0.108
orbitrap	827.424	2	y16	-0.045
orbitrap	1008.51	1	b10	0.133
orbitrap	1062.54	2	y0	0.055
orbitrap	1653.84	1	y16	-0.111
orbitrap	1722.87	1	y17	-0.486

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. HCD spectrum of GSQISLLICEYSSLSVTLCTP (+2) with annotations from *S. albus* J1074 (4 Dehyd).



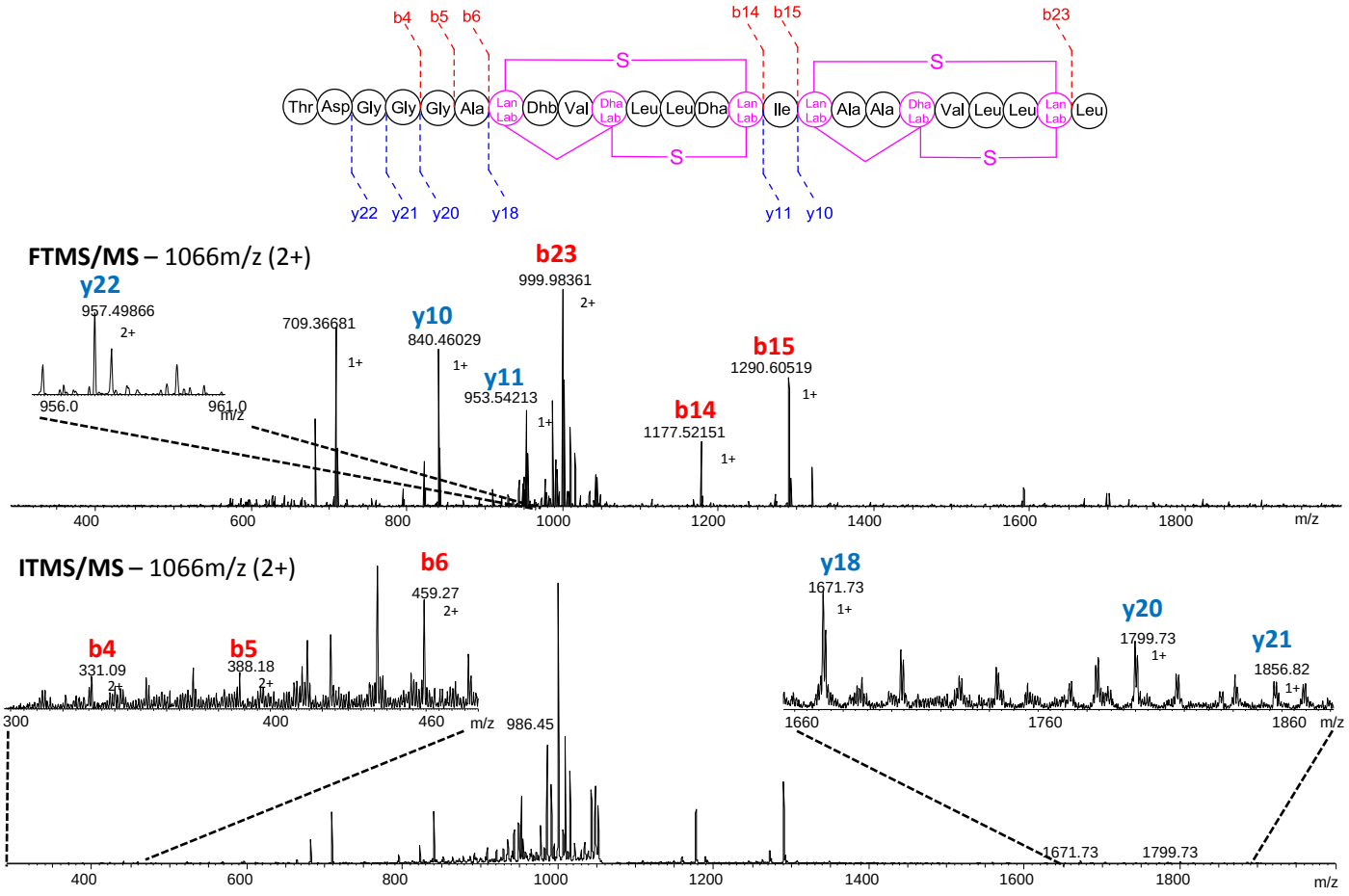
Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
orbitrap	116.07	1	y1	0
orbitrap	199.106	1	y2	-0.004
orbitrap	268.13	1	y3	-0.017
orbitrap	684.41	1	y7	0.057

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. HCD spectrum of TGSQISLLICEYSSLSVTLCTP (+2) with annotations from *S. albus* J1074 (4 Dehyd).



Detector	Observed m/z [Th]	Charge	Species	Difference [Th]
orbitrap	116.071	1	y1	0
orbitrap	141.077	1	b2	0.135
orbitrap	199.113	1	y2	-0.004
orbitrap	268.147	1	y3	-0.018
orbitrap	701.357	1	b7	-0.07
orbitrap	1229.63	1	y12	-0.247

Figure S4. Annotation of lanthipeptide MS/MS spectra discovered by automated peptidogenomics from **Table S2**. HCD spectrum of TGSQISLLICEYSSLSVTLCTP (+2) with annotations from *S. albus* J1074 (5 Dehyd).



Detector	m(obs) [Da]	m(calc) [Da]	Δm [Da]	Δm [ppm]	Fragment
IT	331.09	331.125	-0.035		b4
IT	388.19	388.146	0.044		b5
IT	459.27	459.183	0.087		b6
FT	709.367	709.37	-0.003	4.2	b[S16-C23]
FT	840.46	840.465	-0.005	5.9	y10
FT	953.542	953.549	-0.007	7.3	y11
FT	1177.522	1177.531	-0.009	7.6	b14

Detector	m(obs) [Da]	m(calc) [Da]	Δm [Da]	Δm [ppm]	Fragment
FT	1290.605	1290.615	-0.01	7.7	b15
IT	1671.73	1671.896	-0.166		y18
IT	1799.73	1799.955	-0.225		y20
IT	1856.82	1856.976	-0.156		y21
IT	1897.9	1897.903	-0.003		b[D2-C23]
FT	1913.99	1913.998	-0.008	4.2	y22
FT	1998.96	1998.978	-0.018	9.0	b23
FT	2029.011	2029.025	-0.014	6.9	y23

Figure S5. MS/MS analysis of informatipeptin.

Organism	Precursor peptide sequence	Modifications of mature peptide
S. viridochromogenes	**MALLDLQTIETEER----------TDGGGASTVS-LLSCI----SAASVLLCL**	
S. chartreusis	MALLDLQTIESEER----------TDGGGASTVS-LLSCI----SAASVLLCL	N/A
S. avermitilis (Avi)	MALLDLQTMESDEH----------TGGGGASTVS-LLSCV----SAASVLLCL	1Lan + 1Lab
S. cattleya	MALLDLQNMESEEL----------NGGGASTVS-LLSCV----SAGSVILCV	N/A
S. scabiei	MALLDLQTMEADET----------TGTGGPSSLS-VLSCV----SAASITLCL	N/A
S. erythraea (Eri)	MEMVLELQELDAPNELAYG------DPSHGGGSNLSLLASCAN---STVSLLTCH	1Lan + 1Lab
C. acidiphila (AciA)	MTEEMTLLDLQGMEQTETDSWGGS–GHGGGGDSGLSVTG-CNGH--SGISLL-CDL	2Lab
S. coelicolor (SapB)	MNLFDLQSMETPKEEAMGDVE---------TGSRAS-LLLCGD---SSLSITTCN	1Lan + 1Lab
S. griseus (AmfS)	MALLDLQAMDTPAEDSFGELR---------TGSQVS-LLVCEY---SSLSVVLCTP	1Lan + 1Lab
Kribella flavida (FlaA)	MALLDLQGLETPGYGHGGHH---------HGGSTLTVLG-CGSQRPSNLSLLLCH	N/A

Figure S6. Sequence comparison of precursor peptides in homologous class III lanthipeptide gene clusters in *Streptomyces* genomes. Ser/Ser/Cys motifs for lanthionine/labionin posttranslational modification are highlighted in red and orange. Characterized modifications of corresponding lanthipeptides are listed based on Voeller, *et al. Chembiochem* (2012) , Wang, *et al. ACS Chem. Biol*. (2012) and Voeller, *et al. J. Am. Soc. Chem*. (2013). N/A – modifications not known.

A – *Streptomyces roseosporus* NRRL 15998

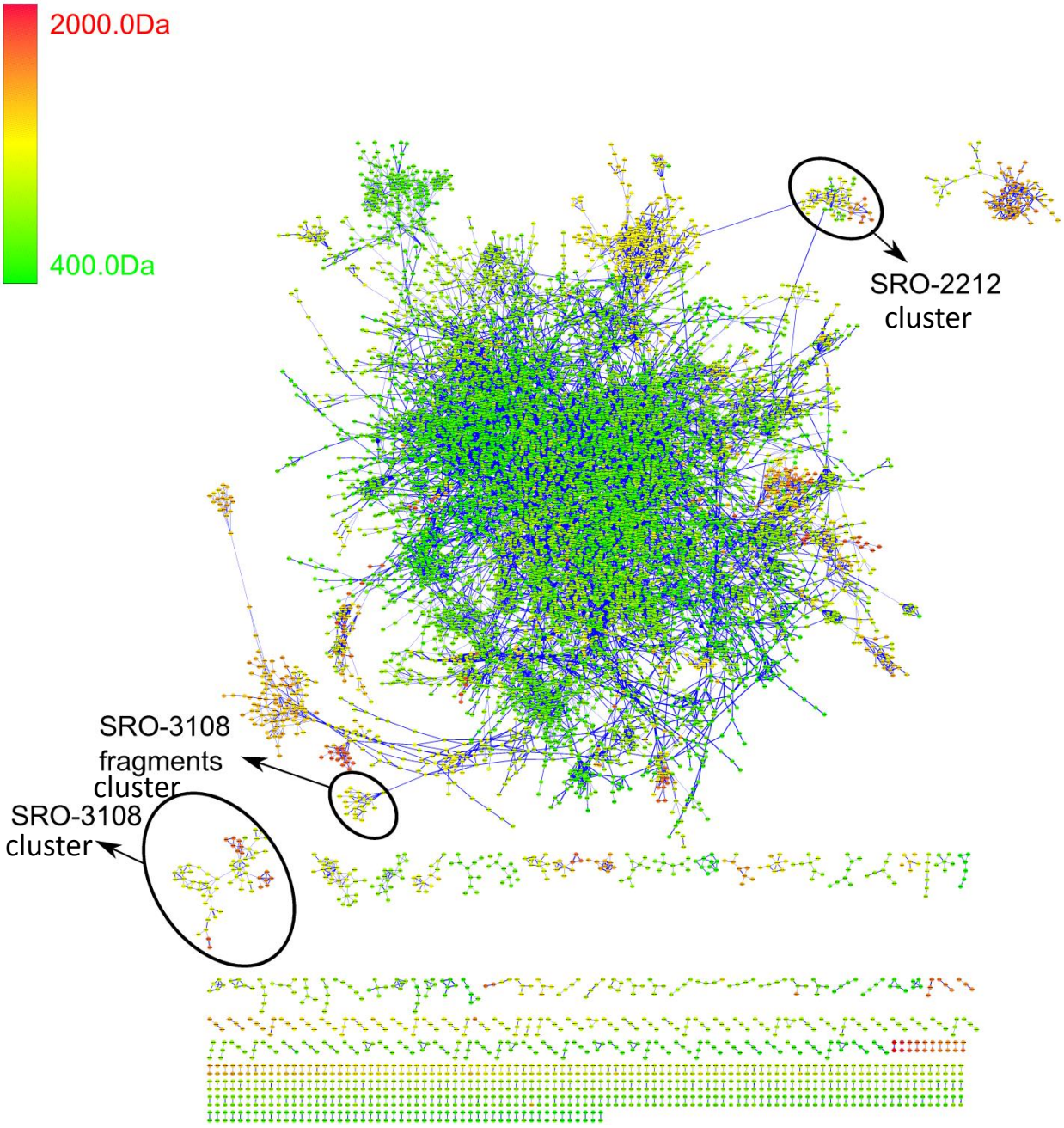
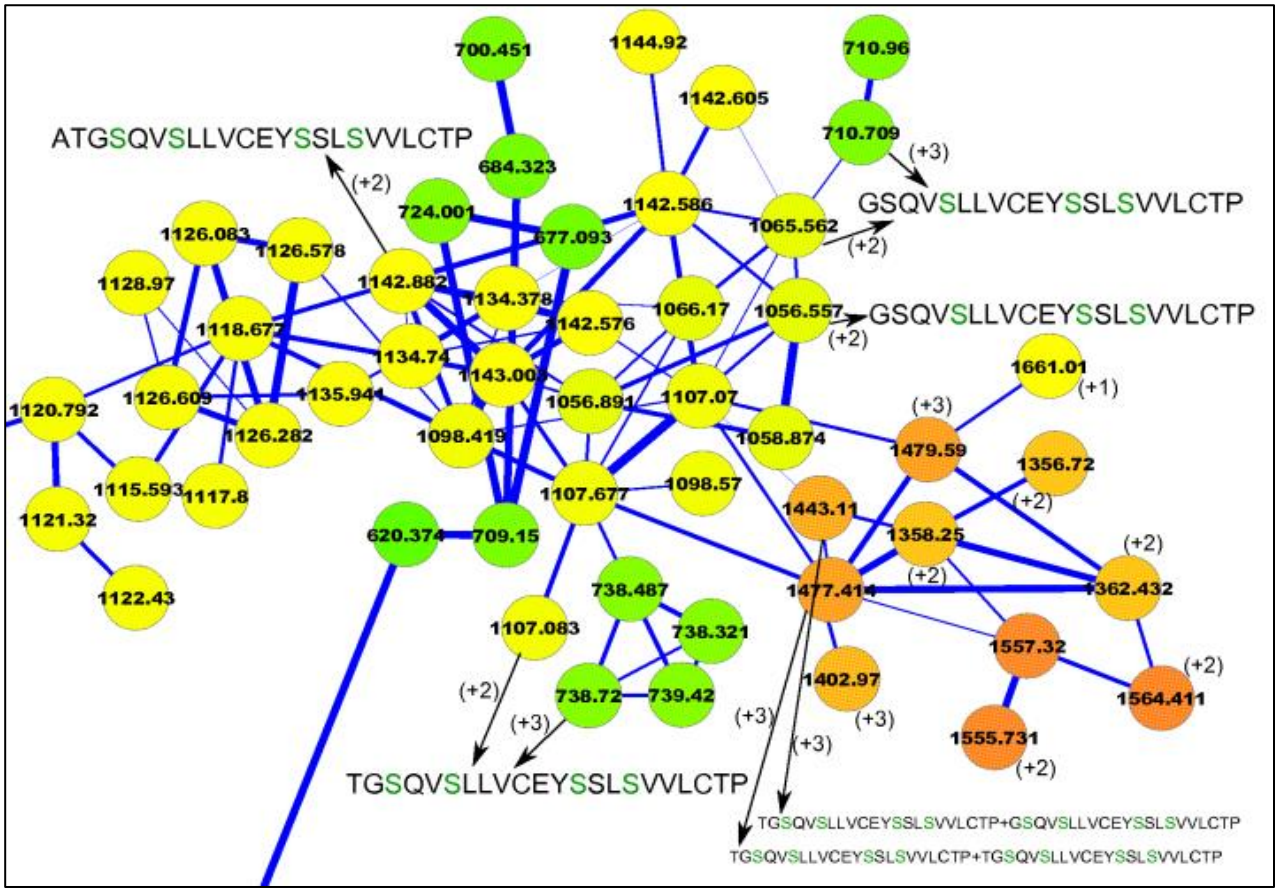


Figure S7. Spectral networks of MS/MS datasets for characterization of lanthipeptide homologs and PSM confirmation. (A) Spectral network of *Streptomyces rososporus* NRRL 15998 with SRO-2212 and SRO-3108 spectral clusters.

SRO-2212 cluster



SRO-3108 fragments cluster

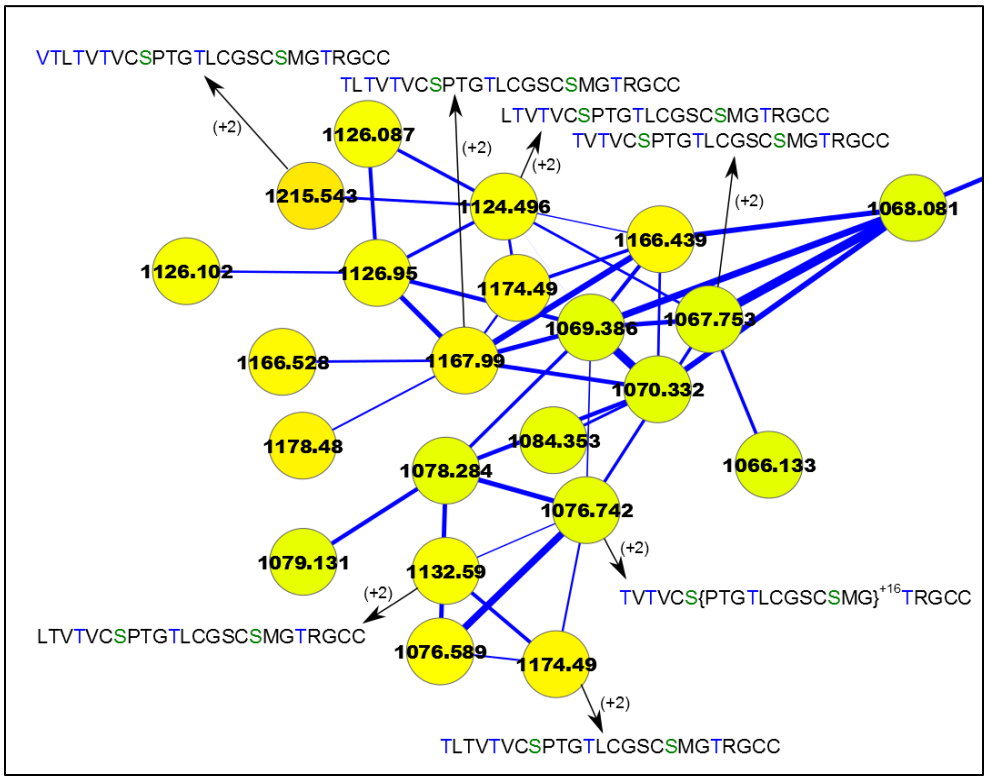


Figure S7. Spectral networks of MS/MS datasets for characterization of lanthipeptide homologs and PSM confirmation. (A) Spectral network of *Streptomyces rososporus* NRRL 15998 with SRO-2212 and SRO-3108 spectral clusters.

SRO-3108 cluster

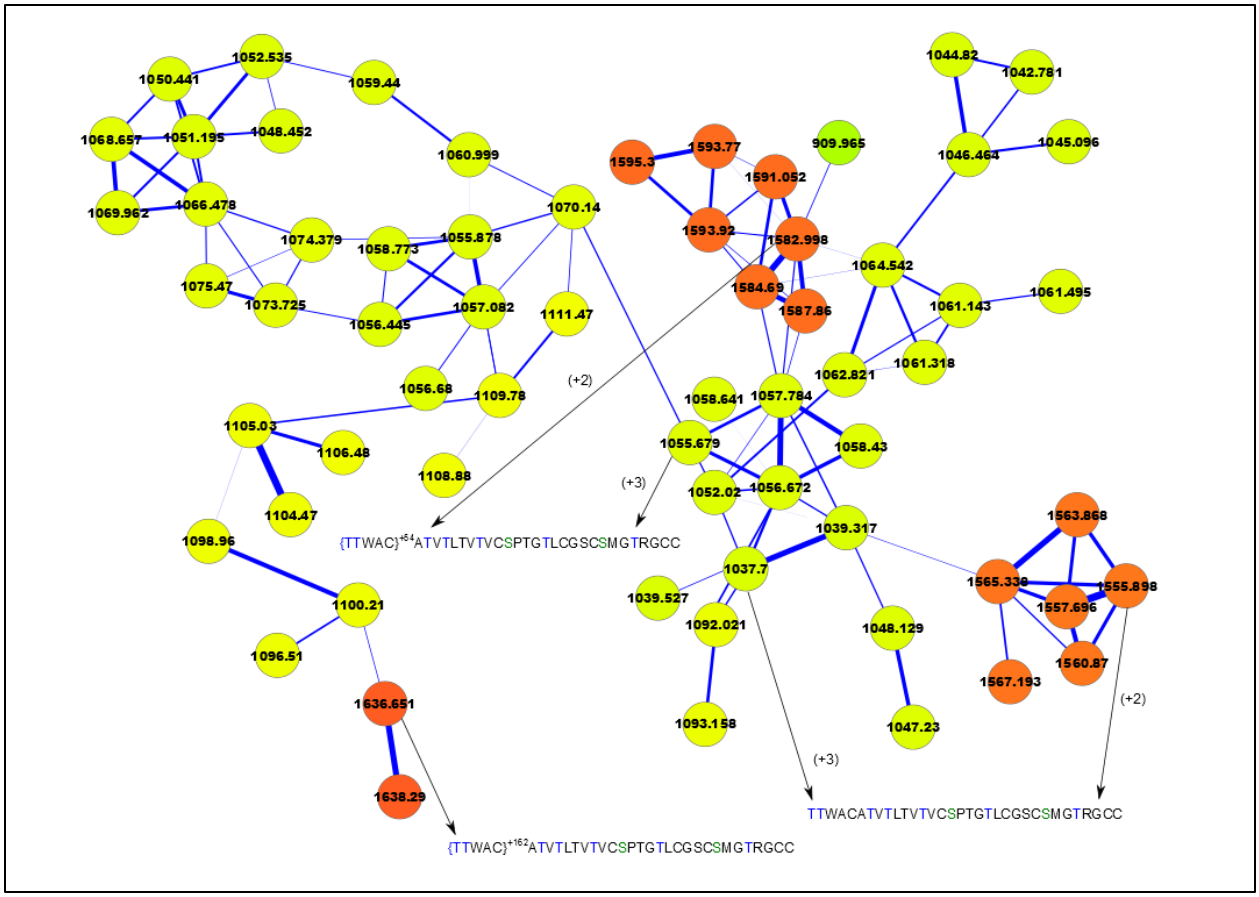


Figure S7. Spectral networks of MS/MS datasets for characterization of lanthipeptide homologs and PSM confirmation. (A) Spectral network of *Streptomyces rososporus* NRRL 15998 with SRO-2212 and SRO-3108 spectral clusters.

B – *Streptomyces viridochromogenes* DSM 40736

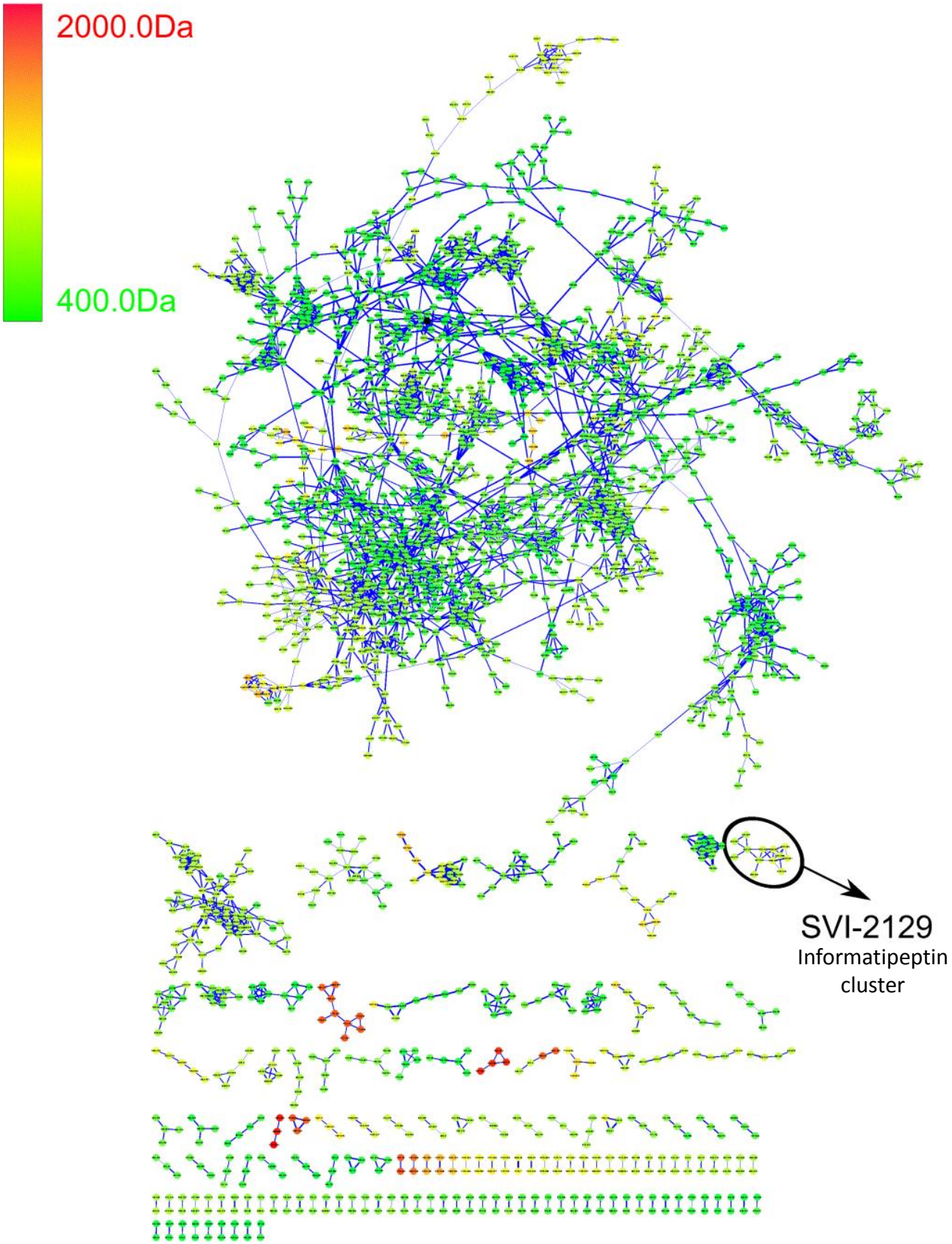


Figure S7. Spectral networks of MS/MS datasets for characterization of lanthipeptide homologs and PSM confirmation. (B) Spectral network of *Streptomyces viridochromogenes* DSM 40736 with the informatipeptin spectral cluster (SVI-2129).

SVI-2129 informatipeptin cluster

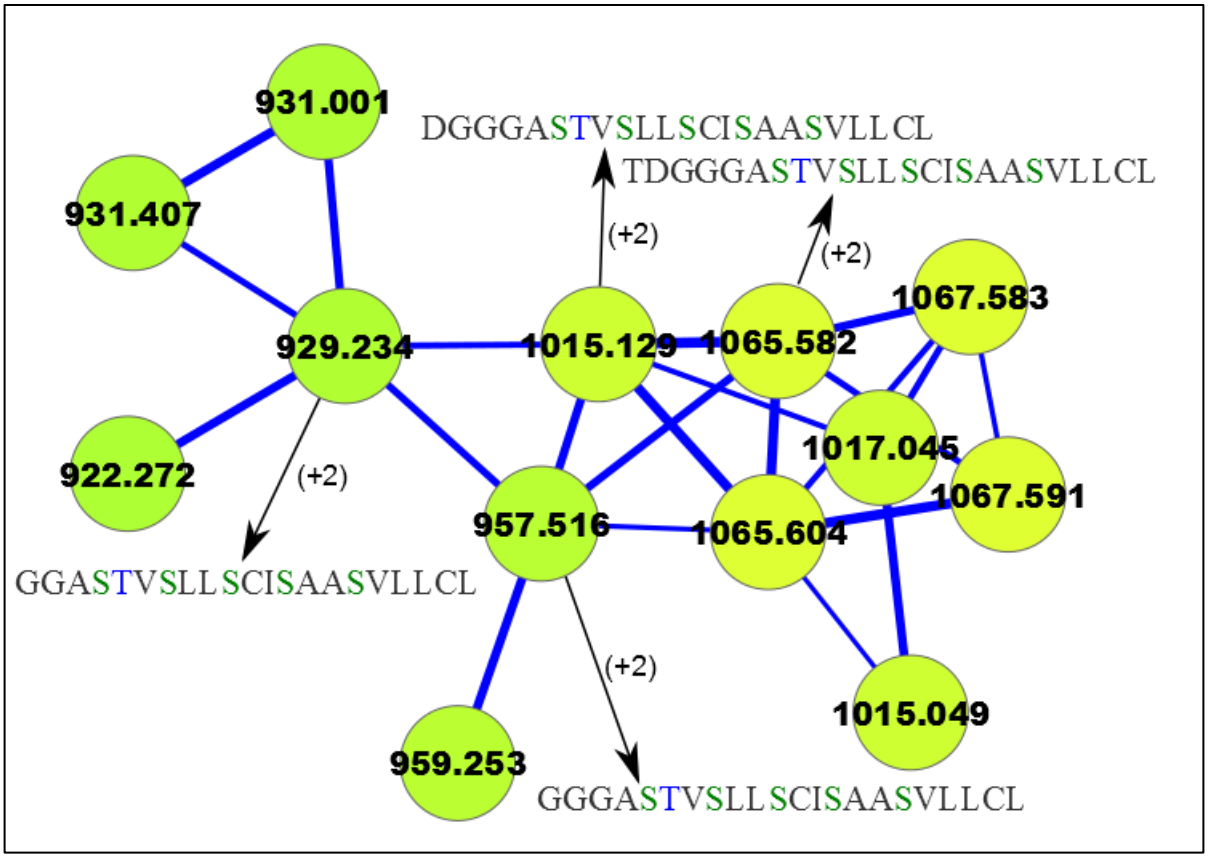


Figure S7. Spectral networks of MS/MS datasets for characterization of lanthipeptide homologs and PSM confirmation. (B) Spectral network of *Streptomyces viridochromogenes* DSM 40736 with the informatipeptin spectral cluster (SVI-2129).

C – *Streptomyces viridochromogenes* DSM 40736

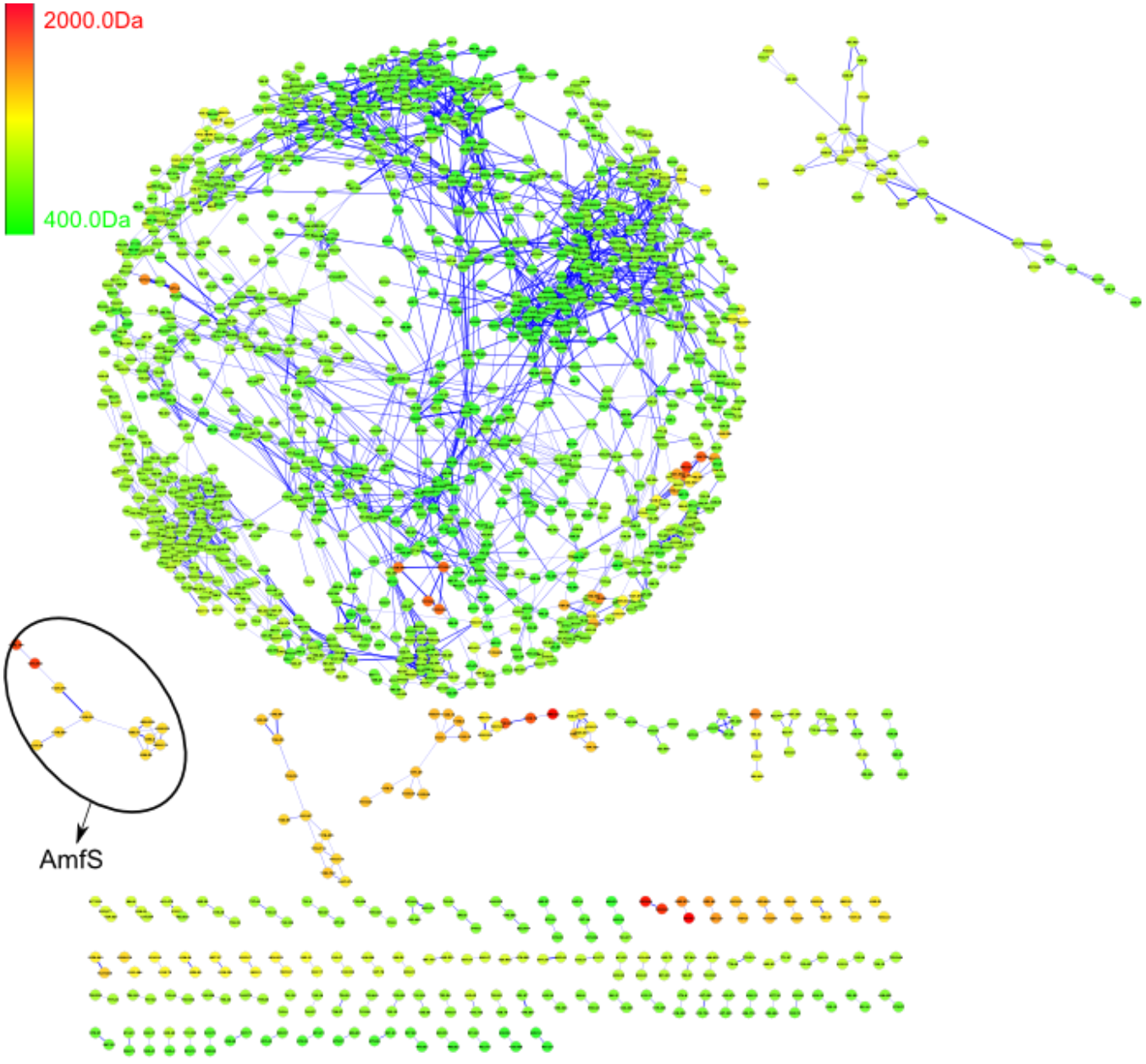


Figure S7. Spectral networks of MS/MS datasets for characterization of lanthipeptide homologs and PSM confirmation. (C) Spectral network of *Streptomyces griseus* IFO 13350 with the AmfS spectral cluster.

AmfS cluster

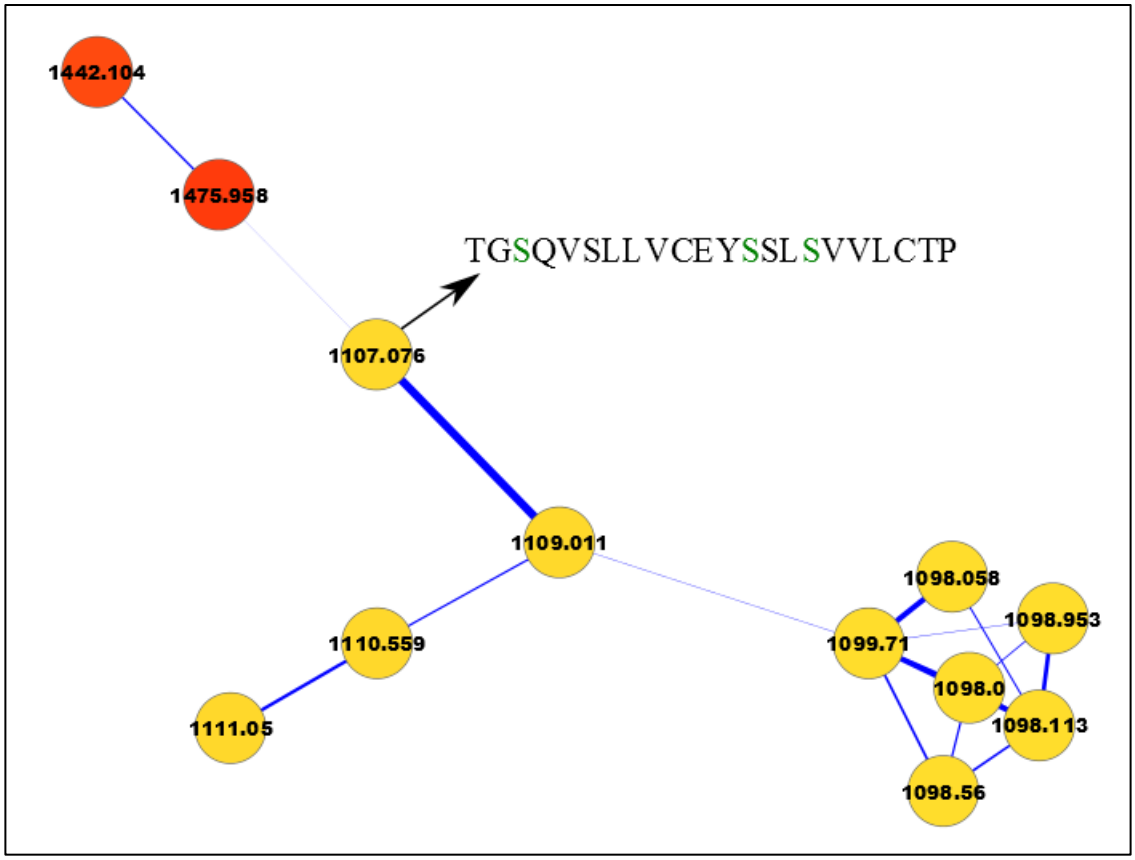


Figure S4. Spectral networks of MS/MS datasets for characterization of lanthipeptide homologs and PSM confirmation. (C) Spectral network of *Streptomyces griseus* IFO 13350 with the AmfS spectral cluster.

D – *Streptomyces albus* J1074

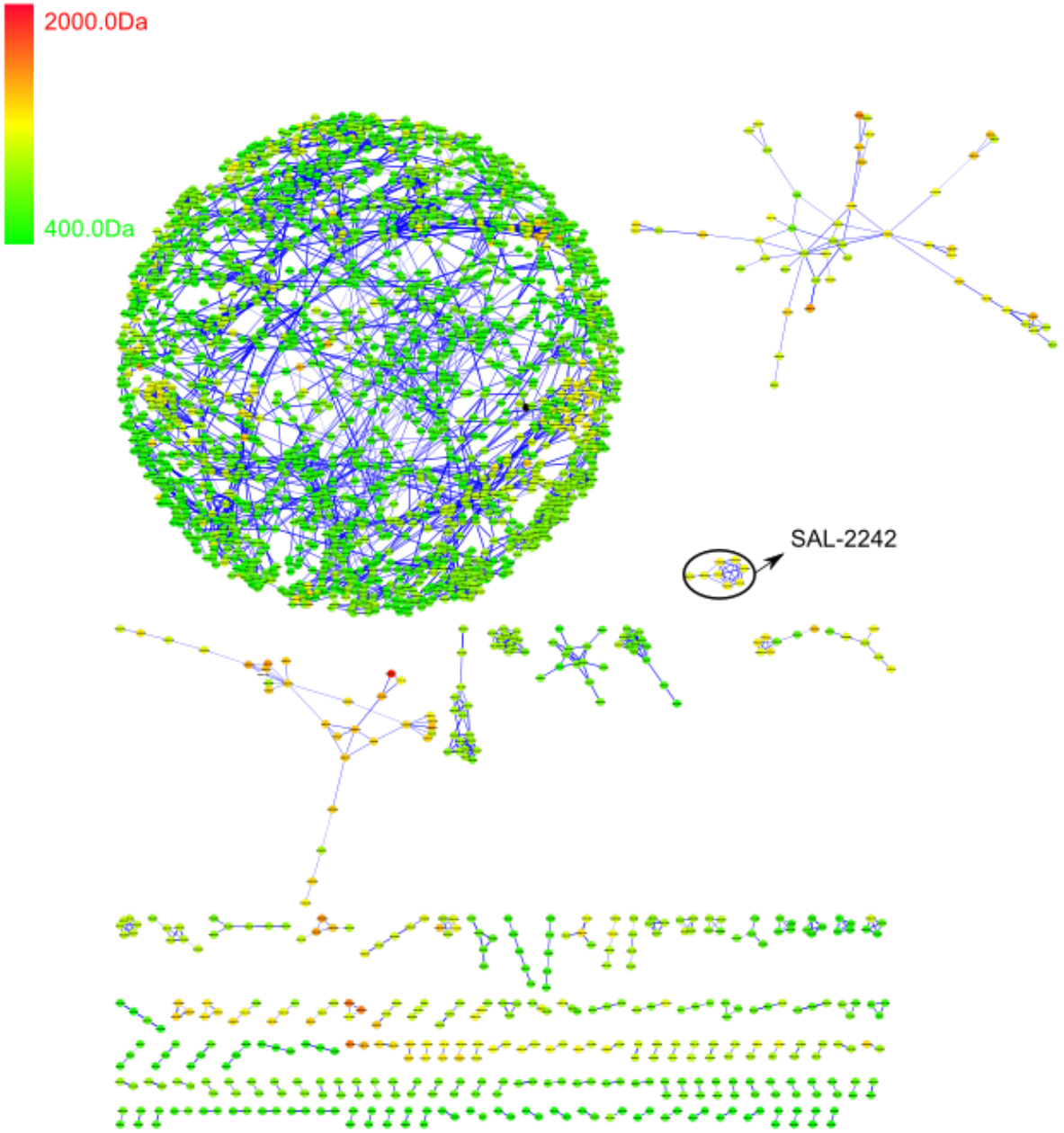


Figure S7. Spectral networks of MS/MS datasets for characterization of lanthipeptide homologs and PSM confirmation. (D) Spectral network of *Streptomyces albus* J1074 with the SAL-2242 spectral cluster.

SAL-2242 cluster

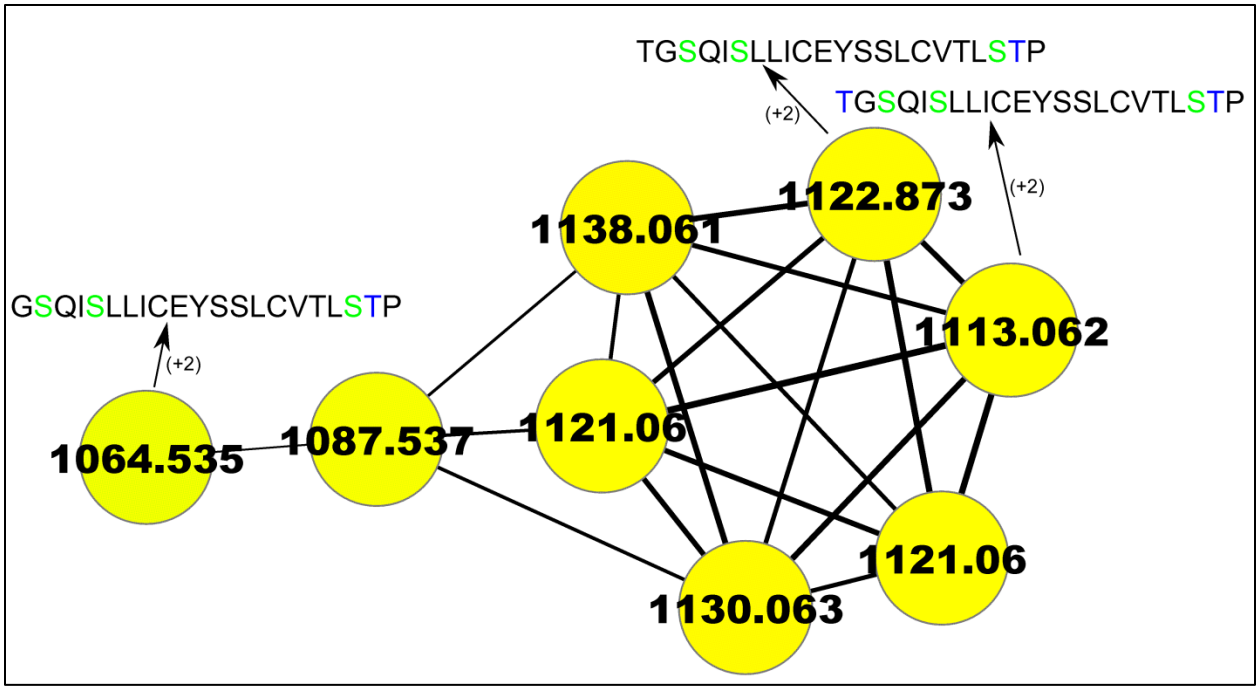


Figure S7. Spectral networks of MS/MS datasets for characterization of lanthipeptide homologs and PSM confirmation. (D) Spectral network of *Streptomyces albus* J1074 with the SAL-2242 spectral cluster.

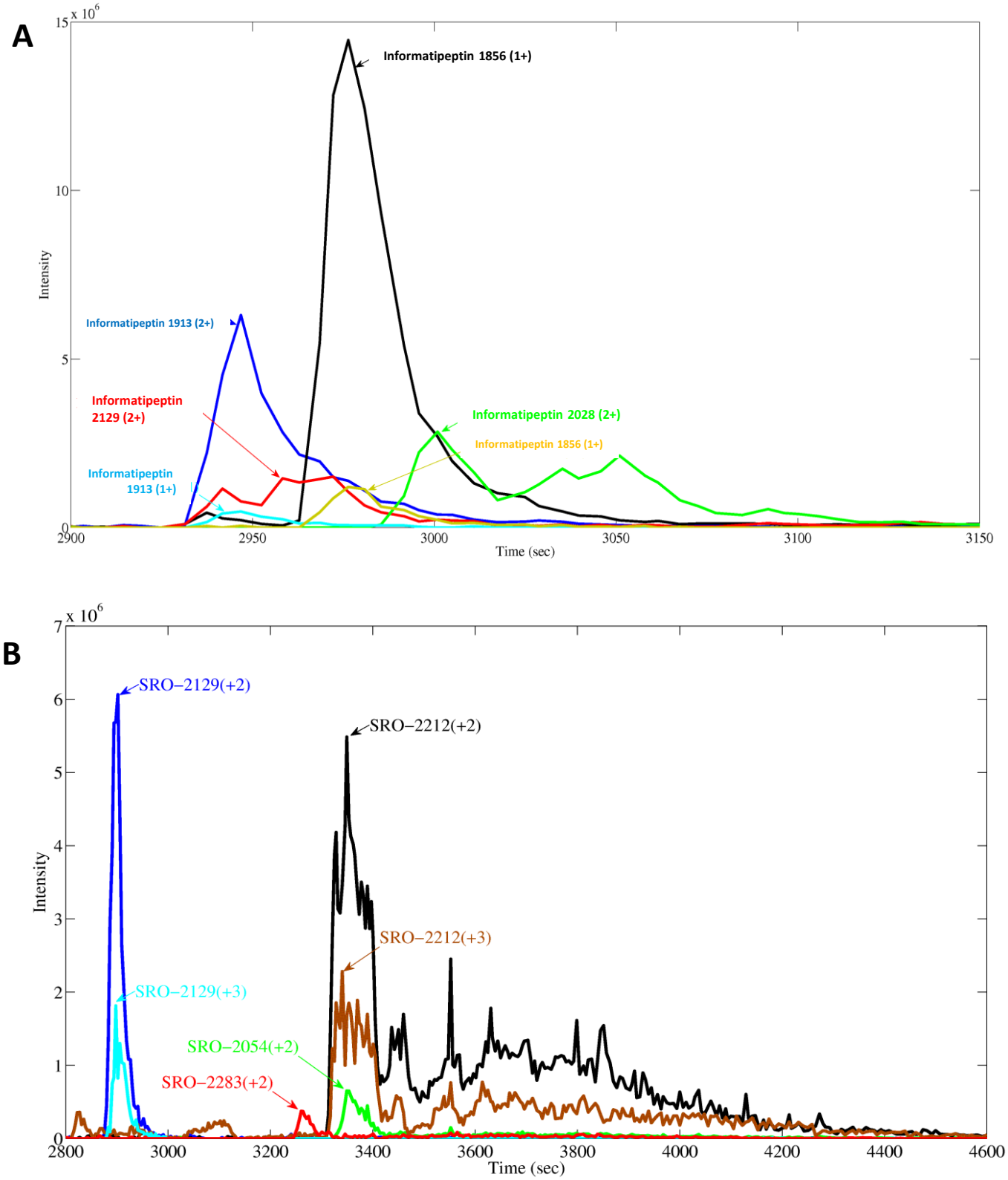


Figure S8. Elution profiles of compounds related to (A) Informatipeptin and (B) SRO-2212. The distinct elution profiles show these compounds are distinct compounds, rather than mass spectrometry adducts.

C

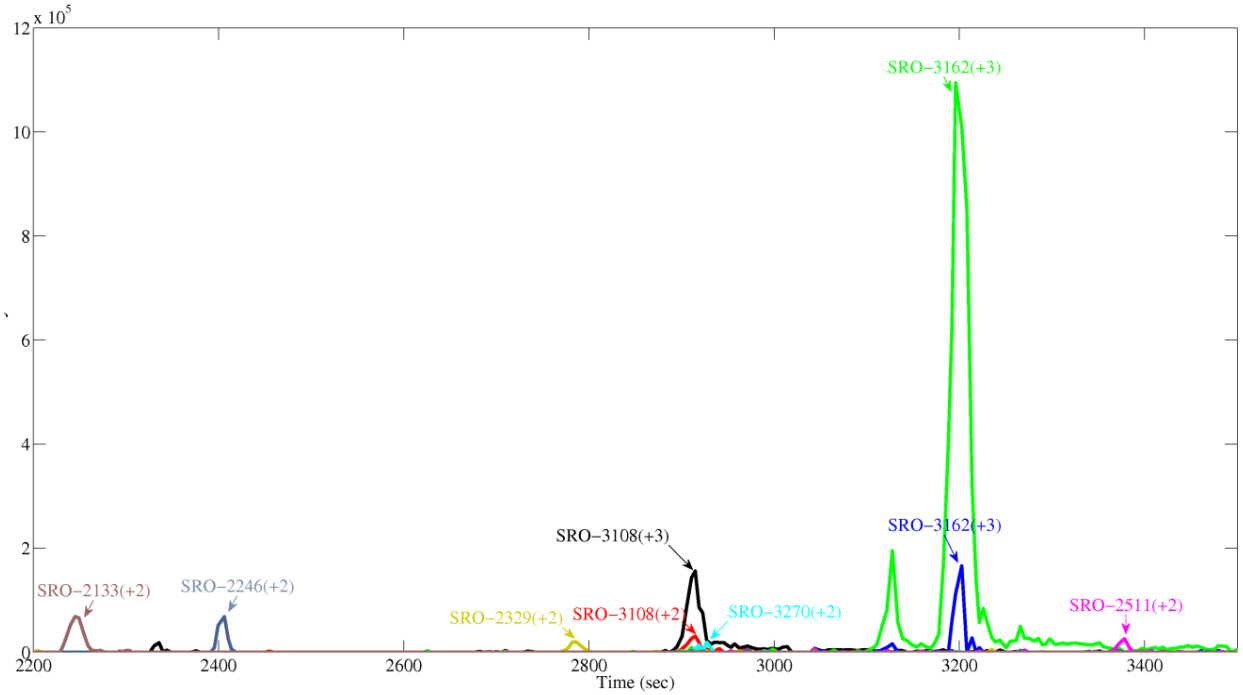


Figure S8. Elution profiles of compounds related to (C) SRO-3108. The distinct elution profiles show these compounds are distinct compounds, rather than mass spectrometry adducts.

Table S1. List of the *Streptomyces* strains used in this study, along with the number of lanthipeptide gene clusters predicted and their known lanthipeptides. Predicted gene clusters are shown in **Fig. S3**.

Species	# Predicted lanthipeptide clusters	Known lanthipeptides
Streptomyces roseosporus NRRL 11379	3	
Streptomyces roseosporus NRRL 15998	2	SRO-2212, SRO-3108
Streptomyces sp. AA4	6	
Streptomyces albus J1074	3	SAL-2242
Streptomyces sviceus ATCC 29083	1	
Streptomyces ghanaensis ATCC 14672	2	
Streptomyces hygroscopicus ATCC 53653	1	
Streptomyces pristinispiralis ATCC 25486	1	
Streptomyces griseoflavus Tü4000	5	
Streptomyces griseus IFO 13350	6	AmfS
Streptomyces lividans TK24	3	SapB
Streptomyces sp. E14	2	
Streptomyces sp. SPB74	2	
Streptomyces sp. SPB78	1	
Streptomyces viridochromogenes DSM 40736	3	Informatipeptin
Streptomyces sp. Mg1	5	
Streptomyces coelicolor A3(2)	3	SapB

Table S2. Informatipeptin gene cluster analysis from *Streptomyces viridochromogenes* DSM 40736.

Gene	Size [aa]	Predicted function	Closest homolog (similarity/identity) [%/%]
SSQG_07264	470	Protease	protease [Streptomyces chartreusis NRRL 12338] (99/95)
SSQG_07265	714	Sporulation regulator	hypothetical protein SchaN1_19165 [Streptomyces chartreusis NRRL 12338] (96/92)
SSQG_07266	158	Regulatory protein	two-component system response regulator [Streptomyces chartreusis NRRL 12338] (97/93)
SSQG_07267	699	Transporter (RamB)	ABC transporter ATP-binding protein [Streptomyces chartreusis NRRL 12338] (78/74)
N/A	38	Precursor peptide	AmfS protein [Streptomyces chartreusis NRRL 12338] (100/97)
sequence gap	522	Transporter (RamA)	ABC transporter ATP-binding protein [Streptomyces chartreusis NRRL 12338] (92/89)
SSQG_07268	897	Lanthionine synthetase	membrane translocator [Streptomyces chartreusis NRRL 12338] (96/93)
SSQG_07269	842	Regulatory protein	regulatory protein [Streptomyces chartreusis NRRL 12338] (94/91)

References (Supporting Information)

1. Actinomycetales group Database, Broad Institute of Harvard and MIT. http://www.broadinstitute.org/.
2. Livesay, E.A.; Tang, K.; Taylor, B.K.; Buschbach, M.A.; Hopkins, D.F.; LaMarche, B.L.; Zhao, R.; Shen, Y.; Orton, D.J.; Moore, R.J.; Kelly, R.T.; Udseth, H.R.; Smith, R.D. *Anal. Chem.* **2008**, *80*, 294.
3. Kersten, R. D.; Yang, Y. L.; Cimermancic, P.; Nam, S. J.; Fenical, W.; Fischbach, M. A.; Moore, B. S.; Dorrestein, P. C. *Nat. Chem. Biol*. **2011**, *7*, 794.
4. Frank, A. M.; Pevzner, P. A. *Anal. Chem.* **2005**, *77*, 964.
5. Frank, A. M.; Pesavento, J. J.; Mizzen, C. A.; Kelleher, N. L.; Pevzner, P. A. *Genome Res*. **2008**, *80*, 2499.
6. Pevzner, P. A.; Dancik, V.; Tang, C. L. *J. Comput. Biol*. **2000**, *7*, 777.
7. Pevzner, P. A.; Mulyukov, Z.; Dancik, V.; Tang, C. *Genome Res*. **2001**, *11*, 290.
8. Tsur, D.; Tanner, S.; Zandi, E.; Bafna, V.; Pevzner, P. A. *Nat. Biotechnol*. **2005**, *23*, 1562.
9. Kim, S.; Gupta, N.; Pevzner, P. A. J. *Proteome Res*. **2008**, *7*, 3354.
10. Mohimani, H.; Kim, S.; Pevzner, P. A. *Algorithms in Bioinformatics*, **2012**, Springer Berlin Heidelberg.
11. Kim, S.; Pevzner, P. A. *60th American Society for Mass Spectrometry Conference (ASMS)* **2012**.