Supporting Information

A Proteomic Survey of Nonribosomal Peptide and Polyketide Biosynthesis in Actinobacteria

Yunqiu Chen[1,2], Ioanna Ntai[1,2], Kou-San Ju[3], Michelle Unger[1,2], Leonid Zamdborg[4], Sarah J. Robinson[1,2], James R. Doroghazi[3], William W. Metcalf[3], Neil L. Kelleher[*,1-3]

[1]Department of Chemistry, [2]Chemistry of Life Processes Institute, Northwestern University, Evanston, Illinois, 60208, United States
[3]The Institute for Genomic Biology, [4]College of Medicine, University of Illinois at Urbana-Champaign, Urbana, Illinois, 61801, United States
*Corresponding author.
Prof. Neil Kelleher. Phone: +1-847-467-4362   Fax: +1-847-467-3276
E-mail: n-kelleher@northwestern.edu

Supplementary Tables

Supplementary Table 1. Taxonomy of the six actinomycetes yielding NRPS and/or PKS identifications. The taxomony is determined based on the 16S rRNA sequence. Also listed are the growth medium and culturing time where NRPS/PKS were identified by LC-MS/MS.

Strain NRRL #	**Taxonomy**	**Growth medium / time**
F-6133	*Kitasatospora sp.*	4×R2A / day 2
F-6134	*Streptomyces sp.*	4×R2A / day 1
F-6143	*Streptomyces sp.*	4×R2A / day 2
F-6562	*Streptomyces sp.*	4×R2A / day 1
F-6652	*Streptomyces sp.*	MSB / day 1
F-6556	*Nocardiopsis sp.*	4×R2A / day 1

Supplementary Table 2. NRPS/PKS identifications from six Actinobacteria strains using PrISM approach. **(a)** Peptide identifications from OMSSA search for strains F-6562, F-6556 and F-6133, where peptides were identified as coming from different NRPS/PKS proteins. **(b)** Peptide identifications for strains F-6133 and F-6556 from *de novo* sequencing and homology search using PEAKS studio, each peptide with its homologous peptide/protein shown. **(c)** Protein identifications from OMSSA search, where each protein has multiple peptides identified. Peptides listed in Supplementary Table 2(a) do not overlap with those from 2(c).

(a) Peptide identifications from OMSSA

Peptide sequence	**E-value**	**Precursor mass**	**Charge**	**Mass error (Da)**	**Accession #**	**Protein description**
Strain F-6562						
TGAAYLPLDLDHPADR	2.496E-8	576.627	3	0.006	ZP_06907900	non-ribosomal peptide synthetase [Streptomyces pristinaespiralis ATCC 25486]
EIFDPAIASAGGR	5.210E-8	652.338	2	0.005	ZP_06907900	non-ribosomal peptide synthetase [Streptomyces pristinaespiralis ATCC 25486]
GDEALDDLVGFFVNTLVLR	2.087E-3	698.378	3	0.003	ZP_06907900	non-ribosomal peptide synthetase [Streptomyces pristinaespiralis ATCC 25486]
DLTTAYAAR	7.489E-3	491.255	2	0.002	ZP_06907900	non-ribosomal peptide synthetase [Streptomyces pristinaespiralis ATCC 25486]
GVRPQDPVAVLLGR	1.280E-17	492.961	3	0.005	ZP_06271943	amino acid adenylation domain protein [Streptomyces sp. ACTE]
VTGVNLLPSFLAAmPDDR	4.139E-14	968.503	2	0.005	ZP_06271943	amino acid adenylation domain protein [Streptomyces sp. ACTE]
ILTELFALYTAR	4.925E-7	705.906	2	0.005	ZP_06271943	amino acid adenylation domain protein [Streptomyces sp. ACTE]
TVGWFTSEYPVR	8.251E-7	721.362	2	0.006	ZP_06271943	amino acid adenylation domain protein [Streptomyces sp. ACTE]

mGAGTDIPLGAPIAGR	8.387E-12	756.897	2	-0.001	YP_003339147	non-ribosomal peptide synthetase-like protein [Streptosporangium roseum DSM 43021]

Supplementary Table 2(a). Continued

DPAYLAELIR	6.970E-9	580.821	2	0.003	YP_003339147	non-ribosomal peptide synthetase-like protein [Streptosporangium roseum DSM 43021]
ADGNLDFLGR	5.814E-4	539.270	2	0.001	YP_003339147	non-ribosomal peptide synthetase-like protein [Streptosporangium roseum DSM 43021]
GAGIEDVVALALPR	2.278E-5	690.898	2	0.005	YP_001510190	amino acid adenylation domain-containing protein [Frankia sp. EAN1pec]
LAWmQAEYR	2.791E-3	592.285	2	0.001	ZP_04691513	CDA peptide synthetase I [Streptomyces roseosporus NRRL 15998]
RADGAVEYLGR	1.232E-3	603.816	2	0.003	YP_289922	amino acid adenylation [Thermobifida fusca YX]
FVADPFGEPGER	1.528E-7	660.816	2	0.003	YP_001105238	putative non-ribosomal peptide synthetase [Saccharopolyspora erythraea NRRL 2338]
LFDTLFVLR	3.337E-8	562.331	2	0.004	YP_003494015	NRPS/siderophore biosynthesis protein [Streptomyces scabiei 87.22]
VELGEIESVLAR	3.402E-8	657.868	2	0.002	YP_702172	non-ribosomal peptide synthetase [Rhodococcus jostii RHA1]
FVADPYGPPGSR	2.500E-6	631.813	2	0.003	YP_289923	non-ribosomal peptide synthase:amino acid adenylation [Thermobifida fusca YX]
AAG(I/L)TLTPR	2.889E-3	450.259	2	-0.001	YP_003274810	amino acid adenylation domain-containing protein [Gordonia bronchialis DSM 43247]
AEDVVAVALPR	1.761E-9	570.326	2	0.003	ZP_06830500	possible non-ribosomal peptide synthetase [Rhodococcus equi ATCC 33707]
Strain F-6556						
FVADPYGPPGSR	1.949E-5	631.814	2	0.005	YP_289923	non-ribosomal peptide synthase:amino acid adenylation [Thermobifida fusca YX]
TPAMVVNVLPLR	1.983E-5	655.389	2	0.005	YP_289923	non-ribosomal peptide synthase:amino acid adenylation [Thermobifida fusca YX]

DDVVFGTVTSGR	1.678E-4	626.813	2	0.003	YP_003073479	nonribosomal peptide synthetase [Teredinibacter turnerae T7901]
FVADPYGAPGAR	3.772E-6	610.808	2	0.004	ZP_07274839	non-ribosomal peptide synthetase [Streptomyces sp. SPB78]

Supplementary Table 2(a) Continued

GYVGRPDLTAER	9.355E-3	667.346	2	-0.001	ADK54820	non-ribosomal peptide synthase [uncultured soil bacterium
AAADALLR	6.989E-3	400.731	2	-0.008	YP_289921	amino acid adenylation [Thermobifida fusca YX]
QHVVAVALPR	6.883E-7	545.330	2	0.000	ZP_04617272	Non-ribosomal peptide synthetase modules and protein [Yersinia ruckeri ATCC 29473]
THDDYLYSVR	5.778E-3	634.802	2	0.008	ZP_04507696	peptide arylation enzyme [Saccharomonospora viridis DSM 43017]
Strain F-6133						
ADGLLDFLGR	3.117E-6	538.790	2	0.008	ZP_06907900	non-ribosomal peptide synthetase [Streptomyces pristinaespiralis ATCC 25486]
LVEALNPPR	2.729E-5	504.797	2	0.002	YP_003133692	amino acid adenylation enzyme/thioester reductase family protein [Saccharomonospora viridis DSM 43017]
RRPDGALDYLGR	9.508E-5	463.586	3	0.005	ZP_06830731	non-ribosomal peptide synthetase [Rhodococcus equi ATCC 33707]
FVADPFGEPGAR	3.717E-4	631.815	2	0.007	YP_003765295	non-ribosomal peptide synthetase [Amycolatopsis mediterranei U32]
AHGLAITPR	2.167E-3	468.276	2	0.002	YP_002779515	non-ribosomal peptide synthetase [Rhodococcus opacus B4]
DLDAAYTAR	8.559E-6	498.244	2	0.001	ZP_06477171	amino acid adenylation domain protein [Frankia symbiont of Datisca glomerata]
FDLAAPPLLR	4.032E-6	556.829	2	0.004	YP_289922	amino acid adenylation [Thermobifida fusca YX]
LVEVVNPER	5.112E-4	527.799	2	0.001	ADI04487	amino acid adenylation domain-containing protein [Streptomyces bingchenggensis BCW-1]
QDVVFGATVSGR	4.741E-5	618.325	2	0.006	ZP_06587758	non-ribosomal peptide synthetase [Streptomyces roseosporus NRRL 15998]

(b) Peptide identifications from PEAKS Studio

<table>
<tr><th>Peptide from <i>de novo</i> sequencing and its homolog</th><th>PEAKS Spider Score</th><th>Precursor mass</th><th>Charge</th><th>Mass error (Da)</th><th>Protein accession#</th><th>Protein description</th></tr>
<tr><td>Strain F-6133</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>RFDLEAPPLLR
|||| |||| |
RFDLAAPPLFR</td><td>28</td><td>663.883</td><td>2</td><td>0.008</td><td>EDY61723</td><td>non-ribosomal peptide synthetase [Streptomyces pristinaespiralis ATCC 25486]</td></tr>
<tr><td>AELFDPVVASAGGR
 ||||| ||||||
 ELFDPALASAGGR</td><td>21.92</td><td>694.865</td><td>2</td><td>0.008</td><td>EDY61723</td><td>non-ribosomal peptide synthetase [Streptomyces pristinaespiralis ATCC 25486]</td></tr>
<tr><td>QGPEKVVALGLPR
 |||| |||||||
AGPEKLVALGLPR</td><td>22.15</td><td>682.39</td><td>2</td><td>-0.03</td><td>EEN84327</td><td>amino acid adenylation domain protein [Rhodococcus erythropolis SK121]</td></tr>
<tr><td>LDGSVEYLGR
 || ||||||
PDGAVEYLGR</td><td>23</td><td>554.79</td><td>2</td><td>0.011</td><td>CAM02313</td><td>putative non-ribosomal peptide synthetase [Saccharopolyspora erythraea NRRL 2338]</td></tr>
<tr><td>QEVVALGLPR
+ ||||||||
ETVVALGLPR</td><td>23</td><td>541.32</td><td>2</td><td>-0.002</td><td>EEN84327</td><td>amino acid adenylation domain protein [Rhodococcus erythropolis SK121]</td></tr>
<tr><td>SIDmVALSFAVLR
| ||| |||||
SVDMVVALFAVLR</td><td>20.54</td><td>719.39</td><td>2</td><td>-0.002</td><td>EEN84327</td><td>amino acid adenylation domain protein [Rhodococcus erythropolis SK121]</td></tr>
</table>

<table>
<tr><td>AEGLALSPR
| || ||||
AAGLSLSPR</td><td>20</td><td>457.26</td><td>2</td><td>0.004</td><td>EEN84327</td><td>amino acid adenylation domain protein [Rhodococcus erythropolis SK121]</td></tr>
</table>

Supplementary Table 2(b). Continued

<table>
<tr><td>TPAALAAVADTR
|||||| | |
TPAALARAAATA</td><td>20.5</td><td>578.82</td><td>2</td><td>0.003</td><td>YP_001821960</td><td>putative NRPS [Streptomyces griseus subsp. griseus NBRC 13350]</td></tr>
<tr><td>LWTVTALFAVLR
| | |||||||
LEMVVALFAVLR</td><td>16.63</td><td>695.34</td><td>2</td><td>-0.15</td><td>EDY6172</td><td>non-ribosomal peptide synthetase [Streptomyces pristinaespiralis ATCC 25486]</td></tr>
<tr><td>Strain F-6556</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>WDEDGLLDYLGR
| || ||||||
WQADGTLDYLGR</td><td>24.5</td><td>726.347</td><td>2</td><td>0.009</td><td>ACR31310</td><td>Arthrofactin synthetase/syringopeptin synthetase C-related non-ribosomal peptide synthetase [Burkholderia glumae BGR1]</td></tr>
<tr><td>DDGPDEVVTGTQGR
|+| ||||| |
DEGAPHVVTGTAGS</td><td>20.5</td><td>723.332</td><td>2</td><td>0.009</td><td>ABB73286</td><td>putative polyketide synthase [uncultured bacterium 2063G]</td></tr>
<tr><td>TDGTLDFLGR
 |||||+|||
ADGTLDYLGR</td><td>25.5</td><td>547.781</td><td>2</td><td>0.009</td><td>ACR31310</td><td>Arthrofactin synthetase/syringopeptin synthetase C-related non-ribosomal peptide synthetase [Burkholderia glumae BGR1]</td></tr>
<tr><td>HDVVFGQPVSGR
 ||||| ||||
DDVVFGGTVSGR</td><td>22.5</td><td>649.339</td><td>2</td><td>0.008</td><td>EEN85381</td><td>amino acid adenylation domain protein [Rhodococcus erythropolis SK121]</td></tr>
<tr><td>[WW]GPGDTVALALPR
    |||| ||||||+
GL GPGDVVALALPK</td><td>25</td><td>769.914</td><td>2</td><td>0.012</td><td>YP_001106481</td><td>putative non-ribosomal peptide synthetase [Saccharopolyspora erythraea NRRL 2338]</td></tr>
</table>

FVADPFSGGPKR |||||| | | FVADPFGGPGDR	25.79	639.319	2	-0.031	YP_002782356	non-ribosomal peptide synthetase [Rhodococcus opacus B4]

(c) protein identifications from OMSSA

Strain NRRL#	Protein description	Protein accession #	OMSSA Protein score	Sequence coverage(%)	# Peptides identified	Protein size (kDa)
F-6133	modular polyketide synthase BFAS1 [Kitasatospora setae KM-6054]	BAJ33099	-829.74	3.08	9	495.9
	modular polyketide synthase BFAS2 [Kitasatospora setae KM-6054]	BAJ33100	-2072.05	4.9	17	517.3
	modular polyketide synthase BFAS3 [Kitasatospora setae KM-6054]	BAJ33101	-701.79	2.37	7	408.5
	modular polyketide synthase BFAS4 [Kitasatospora setae KM-6054]	BAJ33102	-862.46	2.53	7	354.9
	modular polyketide synthase BFAS5 [Kitasatospora setae KM-6054]	BAJ33099	-524.83	4.22	8	218.0
F-6143	non-ribosomal peptide synthetase [Streptomyces sp. e14]	ZP_06712119	-791.23	29.47	7	40.4
	non-ribosomal peptide synthetase [Streptomyces sp. e14]	ZP_06712118	-718.39	27.46	11	73
	non-ribosomal peptide synthetase [Streptomyces sp. e14]	ZP_0671212	-362.05	20.93	8	63
F-6134	modular polyketide synthase [Streptomyces avermitilis MA-4680]	NP_821590	-255.97	2.98	8	353.4
	modular polyketide synthase [Streptomyces avermitilis MA-4680]	NP_821591	-1264.31	2.83	11	642.0
	modular polyketide synthase [Streptomyces avermitilis MA-4680]	NP_821592	-287.04	3.43	5	192.2
	modular polyketide synthase [Streptomyces avermitilis MA-4680]	NP_821593	-99.99	2.05	6	371.3
	modular polyketide synthase [Streptomyces avermitilis MA-4680]	NP_821594	-308.06	2.10	12	806.3
	peptide synthetase [Streptomyces anulatus]	ADG27359	-69.47	2.17	7	462.2
F-6652	amino acid adenylation domain protein [Streptomyces flavogriseus ATCC 33331]	ZP_05803230	-185.68	4.27	15	511.1
	amino acid adenylation domain protein [Streptomyces flavogriseus ATCC 33331]	ZP_05803229	-407.69	10.53	15	226.2
	amino acid adenylation domain protein [Streptomyces flavogriseus ATCC 33331]	ADW01610	-346.49	2.66	6	339.8

Supplementary Table 3. Primers used for PCR amplification of target region for (**a**) F-6133 (**b**) F-6562 and (**c**) F-6556.

(a) PCR primers for strain F-6133

Peptide Sequence	Primer	Primer Sequence
ADGLLDFLGR	6133 1F	CTGGACTTCCTGGGCCGC
	6133 1DF[1]	MTSMTSGAYTTCMTSGGNMGB[2]
	6133 1R	GCGGCCCAGGAAGTCCAG
	6133 1DR	VCKNCCSAKGAARTCSAKSAK
LVEALNPPR	6133 2F	GAGGCCCTGAACCCGCC
	6133 2DF	GARGCSMTSAACCCSCCS
	6133 2R	GGCGGGTTCAGGGCCTC
	6133 2DR	SGGSGGGTTSAKSGCYTC
RFDLEAPPLLR	6133 3R	GGCCTCCAGGTCGAAGCG
	6133 3DR	GGSGCYTCSAKRTCGAAVCK
AELFDPVVASAGGR	6133 4F	TTCGACCCGGTCGTCGC
	6133 4DF	TTCGAYCCSGTSGTSGC
QGPEKVVALGLPR	6133 5F	GTCGTCGCCCTGGGCCT
	6133 5DF	GTSGTSGCSMTSGGNMTS
	6133 5R	AGGCCCAGGGCGACGAC
	6133 5DR	SAKNCCSAKSGCSACSAC
LDGSVEYLGR	6133 6F	TCCGTCGAGTACCTGGGC
	6133 6DF	GGNWSSGTSGARTAYMTSGG
	6133 6R	GCCCAGGTACTCGACGGA
	6133 6DR	CCSAKRTAYTCSACSSWNCC
EQVVALGLPR	6133 7F	GTCGTCGCCCTGGGCCT
	6133 7DF	GTSGCSMTSGGNMTSCC
SIDMVALSFAVLR	6133 8F1	GACATGGTCGCCCTGTCCTT
	6133 8DF1	GAYATGGTSGCSMTSWSSTTC
	6133 8F2	TCCTTCGCCGTCCTGCG
	6133 8DF2	WSSTTCGCSGTSMTSMGB
	6133 8R1	AAGGACAGGGCGACCATGTC
	6133 8DR1	GAASSWSAKSGCSACCATRT C
	6133 8R2	CGCAGGACGGCGAAGGA
	6133 8DR2	VCKSAKSACSGCGAASSW
EAGLALSPR	6133 9F	GGCCTGGCCCTGTCCCC
	6133 9DF	GGNMTSGCSMTSWSSCC
	6133 9R	GGGGACAGGGCCAGGCC
	6133 9DR	GGSSWSAKSGCSAKNCC
TPAALAAVADTR	6133 10F	CGTCGCCGACACCCGC
	6133 10DF	GCSGTSGCSGAYACSMG
	6133 10R	GCGGGTGTCGGCGACG
	6133 10DR	CKSGTRTCSGCSACSGC

Supplementary Table 3(a). Continued

LWTVTALFAVLR	6133 11F	ACCGCCCTGTTCGCCGT
	6133 11DF	ACSGCSMTSTTCGCSGT
	6133 11R	ACGGCGAACAGGGCGGT
	6133 11DR	ACSGCGAASAKSGCSGT
RRPDGALDYLGR	6133 12F	CTGGACTACCTGGGCCGC
	6133 12DF	GCSMTSGAYTAYMTSGGNMGB
	6133 12R	GCGGCCCAGGTAGTCCAG
	6133 12DR	VCKNCCSAKRTARTCSAKSG C
FVADPFGEPGAR	6133 14F	GCCGACCCGTTCGGCG
	6133 14DF	GCSGAYCCSTTCGGNGAR
	6133 14R	CGCCGAACGGGTCGGC
	6133 14DR	YTCNCCGAAS GGRTCSGC
AGDLAYALATGR	6133 15F	GCCTACGCCCTGGCCAC
	6133 15DF	GCSTAYGCSMTSGCSAC
	6133 15R	GTGGCCAGGGCGTAGGC
	6133 15DR	GTSGCSAKSGCRTASGC
AHGLAITPR	6133 16R	CGCGGGGTGATGGCCA
	6133 16DR	CKSGGSGTSAKTSGCSA
DLDAAYTAR	6133 17F	GACGCCGCCTACACCGC
	6133 17DF	GAYGCSGCSTAYACSGC
	6133 17R	GCGGTGTAGGCGGCGTC
	6133 17DR	GCSGTRTASGCSGCRTC
FDLAAPPLLR	6133 18F	TTCGACCTGGCCGCCC
	6133 18DF	TTYGAYMTSGCSGCSCC
LVEVVNPER	6133 19F	GAGGTCGTCAACCCGGAGC
	6133 19DF	GARGTSGTSAAYCCSGARMG
QDVVFGATVSGR	6133 20F	GTCTTCGGCGCCACCGTC
	6133 20DF	GTSTTCGGNGCSACSGT

(b) PCR primers for strain F-6562

Peptide Sequence	Primer	Primer Sequence
TGAAYLPLDLDHPADR	6562 1DF	GCSGCSTAYMTSCCSMTSGA
	6562 1F	GCCGCCTACCTGCCGCTG
EIFDPAIASAGGR	6562 2DF	GARMTSTTCGAYCCSGCSMT
	6562 2F	GAGATCTTCGACCCGGCCAT
	6562 2DR	AKSGCSGGRTCGAASAKYTC
	6562 2R	ATGGCCGGGTCGAAGATCTC
GDEALDDLVGFFVNTLVLR	6562 3DF	GAYMTSGTSGGNTTYTTYGTS AA
	6562 3F	CTGGTCGGCTTCTTCGTCAAC
DLTTAYAAR	6562 4DR	GCSGCRTASGCSGTSGT
	6562 4R	GCGGCGTAGGCGGTGGT

Supplementary Table 3 (b). Continued

GVRPQDPVAVLLGR	6562 5DF	CCSGTSGCSGTSMTSMT
	6562 5F	CCGGTCGCCGTCCTGCT
VTGVNLLPSFLAAmPDDR	6562 6DF1	CSGCSATGCCSGAYGA
	6562 6DF2	WSSTTYMTSGCSGCSATG
	6562 6F	CCGCCATGCCGGACGA
	6562 6DR1	TCRTCSGGCATSGCSG
	6562 6R	TCGTCCGGCATGGCGG
ILTELFALYTAR	6562 7DR	CKSGCSGTRTASAKSGCRAA
	6562 7R	GGGCGGTGTACAGGGCGAA
TVGWFTSEYPVR	6562 8DF	TGGTTYACSWSSGARTAYCCSG
	6562 8F	TGGTTCACCTCCGAGTACCCG
	6562 8DR	CSGGRTAYTCSSWSGTRAACCA
	6562 8R	CGGGTACTCGGAGGTGAACCA
mGAGTDIPLGAPIAGR	6562 9DF	GGNACSGAYMTSCCSMTSG
	6562 9F	GGCACCGACATCCCGCT
DPAYLAELIR	6562 10DF	GCSTAYMTSGCSGARMTSMT
	6562 10F	GCCTACCTGGCCGAGCTGATC
	6562 10DR	AKSAKYTCSGCSAKRTASGC
	6562 10R	GATCAGCTCGGCCAGGTAGGC
ADGNLDFLGR	6562 11DR	GAARTCSAKRTTNCCRTCSGC
	6562 11R	AAGTCCAGGTTGCCGTCGGC
GAGIEDVVALALPR	6562 12DF	GARGAYGTSGTSGCSMTSGC
	6562 12F	GAGGACGTCGTCGCCCTGG
LAWmQAEYR	6562 13DF	GCSTGGATGCARGCSGARTA
	6562 13F	GCCTGGATGCAGGCCGAGTA
RADGAVEYLGR	6562 14DR	AKRTAYTCSACSGCNCCRTC
	6562 14R	GGTACTCGACGGCGCCGTC
FVADPFGEPGER	6562 15DR	GYTCNCCRAASGGRTCSGC
	6562 15R	TCGCCGAACGGGTCGG
LFDTLFVLR	6562 16DF	TTYGAYACSMTSTTYGTSMTSCG
	6562 16F	TTCGACACCCTGTTCGTCCTG
	6562 16DR	CGSAKSACRAASAKSGTRTC RAA
	6562 16R	CAGGACGAACAGGGTGTCGA A
VELGEIESVLAR	6562 17DR	TCSAKYTCNCCSAKYTCSAC
	6562 17R	TCGATCTCGCCCAGCTCGAC
FVADPYGPPGSR	6562 18DR	CCRTASGGRTCSGCSACRAA
	6562 18R	GTACGGGTCGGCGACGAA

Supplementary Table 3 (b). Continued

AAG(I/L)TLTPR	6562 19DR	GSGTSAKSGTSAKNCCSGC
	6562 19R	GCSGGNMTSACSMTSACSC
ADGAVEYIGR	6562 20DR	RTAYTCSACSGCNCCRTCSG
	6562 20R	GTACTCGACGGCGCCGTC

(c) PCR Primers for strain F-6556

Peptide Sequence	Primer	Primer Sequence
TPAMVVNVLPLR	6556 1F	ATGGTGGTCAACGTGCTGCC
	6556 1R	GGCAGCACGTTGACCACCAT
	6556 1DF	ATGGTSGTSAAYGTSMTSCCS
	6556 1DR	SGGSAKSACRTTSACSACCAT
FVADPYGPPGSR	6556 2F	GCCGACCCGTACGGCCC
	6556 2R	GGGCCGTACGGGTCGGC
	6556 2DF	GCSGAYCCSTAYGGNCCS
	6556 2DR	SGGNCCRTASGGRTCSGC
FVADPYGAPGAR	6556 3F	GCCGACCCGTACGGCG
	6556 3R	CGCCGTACGGGTCGGC
	6556 3DF	GCNGAYCCSTAYGGNGCNC
	6556 3DR	GNGCNCCRTASGGRTCNGC
THDDYLYSVR	6556 4F	ACCCACGACGACTACCTGTACTCG
	6556 4R	CGAGTACAGGTAGTCGTCGTGGGT
	6556 4DF	ACSCAYGAYGAYTAYMTSTAYWSBGT
	6556 4DR	CVSWRTASAKRTARTCRTCRTGSGT
GYVGRPDLTAER	6556 5F	GACCTGACCGCCGAGCG
	6556 5R	CGCTCGGCGGTCAGGTC
VLVGGEAVPESLAAR	6556 6F	GAATCCCTGGCCGCCCG
	6556 6R	CGGGCGGCCAGGGATTC
FVADPFSGGPKR	6556 7F	GATCCCTTCTCGGGCGGG
	6556 7R	CCCGCCCGAGAAGGGATC
WDEDGLLDYLGR	6556 9F	GCTGGACTACCTGGGCCGC
	6556 9R	GCGGCCCAGGTAGTCCAGC
	6556 9DF	TSMTSGAYTAYMTSGGNCGS
	6556 9DR	SCGNCCSAKRTARTCSAKSA
DDGPDEVVTGTQGR	6556 10F	GTGGTGACCGGCACCCAG
	6556 10R	CTGGGTGCCGGTCACCAC
	6556 10DF	GARGTNGTNACSGGNACSCAR
	6556 10DR	YTGSGTNCCSGTNACNACYTC
TDGTLDFLGR	6556 11F	CTGGACTTCCTGGGCCGC
	6556 11R	GCGGCCCAGGAAGTCCAG
	6556 11DF	CSMTSGAYTTYMTSGGNCGS
	6556 11DR	SCGNCCSAKRAARTCSAKSG

Supplementary Table 3(c). Continued

HDVVFGQPVSGR	6556 12F	GACGTGGTGTTCGGCCAGC
	6556 12R	GCTGGCCGAACACCACGTC
	6556 12DF	GAYGTNGTNTTYGGNCARCCS
	6556 12DR	SGGYTGNCCRAANACNACRTC
[WW]GPGDTVALALPR	6556 13F	GCGACACCGTGGCCCTG
	6556 13R	CAGGGCCACGGTGTCGC
	6556 13DF	GGNCCSGGNGAYACSGT
	6556 13DR	ACSGTRTCNCCSGGNCC

[1] Primers designed based on degenerate codons have a “D” designation, such as 1DF or 1DR, while those based on most likely codons do not.

[2] Standard abbreviations are used: Y, C or T; R, A or G; S, G or C; W, A or T; K, T or G; M, A or C; B, C, G or T; V, A, C or G; N, A, C, T or G.

Supplementary Table 4. PCR reactions completed for analysis of the genome of strains F-6133, F-6562 and F-6556

PCR reaction number	Forward Primer	Reverse Primer	Expected Length (bp)
F-6133			
1[1]	A3F	1R	864
2	4F	A7R	708
3	5F	A7R	1000
4	A3F	6R	750
5	8F1	A7R	975
6	8F2	A7R	975
7	A7F	10R	600
8	11F	A7R	1000
9	A3F	12R	723
10	A3F	13R	609
11	A3F	14R	651
12	KS1	15R	500
14	A7F	16R	555
15	19F	A3R	846
16	KS1	21R	200
17	2F	A7R	1581
18	7F	A7R	1072
19	A3F	9R	1230
20	A3F	16R	1260
21	17F	A3R	1419
22	18F	A3R	1452
23	19F	A7R	1600
24	20F	A3R	1800

Supplementary Table 4. Continued

25	21F	M6R	1200
26	A7F	2R	2715
27	A7F	3R	2400
28	A7F	8R1	2250
29	A7F	8R2	2250
30	12F	A3R	2500
31	A7F	17R	2781
32	17F	A7R	2100
33	18F	A7R	2160
34	20F	A7R	2500
35	1F	A7R	3192
36	A3F	2R	3417
37	A3F	3R	3000
38	4F	A3R	3240
39	A7F	5R	3600
40	6F	A7R	3090
41	9F	A3R	3240
42	10F	A3R	3030
43	A7F	11R	3648
44	A7F	13R	3102
45	14F	A7R	3100
46	A7F	22R	3120
47	22F	A3R	unknown
48	12F	13R	4758
49	12DF	13R	4758
F-6562			
1	1F	A3R	300
2	1F	A7R	1100
3	2F	A7R	720
4	1F	2R	400
5	1DF	2R	400
6	3F	A3R	1000
7	3F	A7R	1700
8	A7F	4R	1100
9	A3F	4R	1800
10	5F	A3R	300
11	5F	A7R	1000
12	6F	A7R	400
13	A3F	6R	300
14	5F	6R	600
15	A7F	7R	2300
16	A7F	8R	1440

Supplementary Table 4. Continued

17	8F	A3R	2300
18	9F	A3R	500
19	9F	A7R	1800
20	A3F	10R	250
21	10F	A7R	486
22	A3F	11R	760
23	12F	A3R	300
24	12F	A7R	1000
25	13F	A7R	630
26	A3F	14R	762
27	A3F	15R	670
28	2F	15R	600
29	16F	A3R	780
30	16F	A7R	1500
31	A7F	16R	1700
32	A3F	17R	762
33	A3F	18R	690
34	10F	18R	450
35	A3F	19R	1250
36	A7F	19R	500
37	A3F	20R	774
F-6556			
1	1F	2R1	1650
2	1F	2R2	1650
3	1F	A3R	950
4	1F	A7R	1650
5	14F	6R	1350
6	A7F	6R	2300
7	A3F	6R	3000
8	12F	13R	700
9	12F	A7R	1650
10	12F	A3R	1000
11	9F	A3R	4400
12	A3F	9R	750
13	9F	A7R	3000
14	4F	10R	300
15	7F	4R	3500
16	6F	4R	3500
17	A3F	2R	750
18	A3F	3R	750
19	5F	12R	2400
20	5F	A3R	3000

Supplementary Table 4. Continued

21	5F	A7R	3700
22	5F	13R	3000
23	A3F	11R	750

[1] Same PCR reactions were set up for primers based on both the most likely codons and degenerate codons. For example, reaction 1 is performed twice, one using A3F and 1R, and the other using A3F and 1DR.

Supplementary Table 5. PCR amplicons from strains NRRL F-6133, F-6562 and F-6556 whose translated nucleotides were identified by OMSSA search. The primers used for the PCR reaction, the length of the translated nucleotides, and the number of peptides detected from them were listed. The most homologous protein is determined by BLASTX analysis of the translated nuleotides of the PCR amplicon.

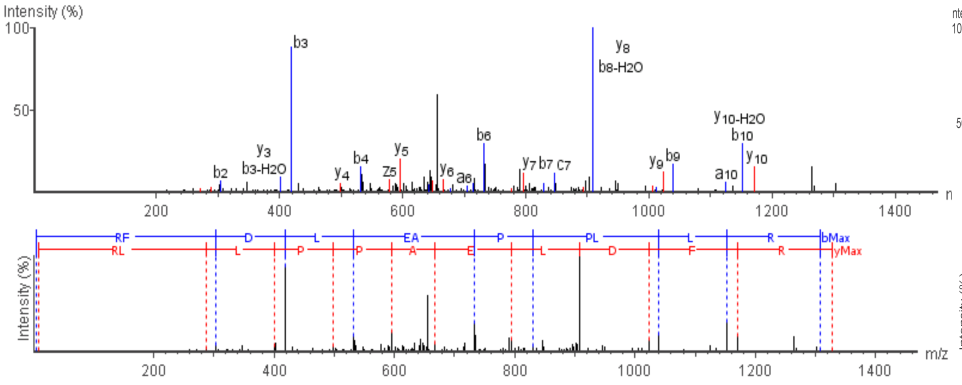
Strain NRRL#	**primers used in PCR reaction**	**Length of translated nucleotides (aa)**	**# Peptides detected**	**Accession # and description of the most homologous protein**	**% identity/similarity at the protein level**
F-6133	6133 2F A7R	140	7	ACZ86404, non-ribosomal peptide synthetase-like protein [*Streptosporangium roseum* DSM 43021]	45/57
	11DF A7R	342	6	EFL03209, non-ribosomal peptide synthetase [*Streptomyces sp.* SPB78]	71/82
F-6562	6562 2F A7R	203	2	ZP_06907900, non-ribosomal peptide synthetase [*Streptomyces pristinaespiralis* ATCC 25486]	78/82
	6562 1DF 6562 2R	108	2	ZP_06907900, non-ribosomal peptide synthetase [*Streptomyces pristinaespiralis* ATCC 25486]	81/88
	6562 2F 6562 15DR	191	1	YP_003339147, non-ribosomal peptide synthetase-like protein [*Streptosporangium roseum* DSM 43021]	77/82
	6562 6F A7R	122	1	ZP_06271943, amino acid adenylation domain protein [*Streptomyces sp.* SA3_actE]	60/65
	6562 9DF A7R	81	1	ZP_06774222, non-ribosimal peptide synthetase [*Streptomyces clavuligerus* ATCC 27064]	63/73
F-6556	6556 1F 6556 2R	472	12	YP_289923, non-ribosomal peptide synthase:amino acid adenylation [*Thermobifida fusca* YX]	50/63
	~~A3F~~	~~181~~	~~5~~	~~YP_289923, non-ribosomal peptide~~	~~63/72~~

	6556 6SR			synthase:amino acid adenylation [*Thermobifida fusca* YX]	
	6556 4SF 6556 10DR	155	2	YP_289927, 2,3-dihydroxybenzoate-AMP ligase [*Thermobifida fusca* YX]	70/75

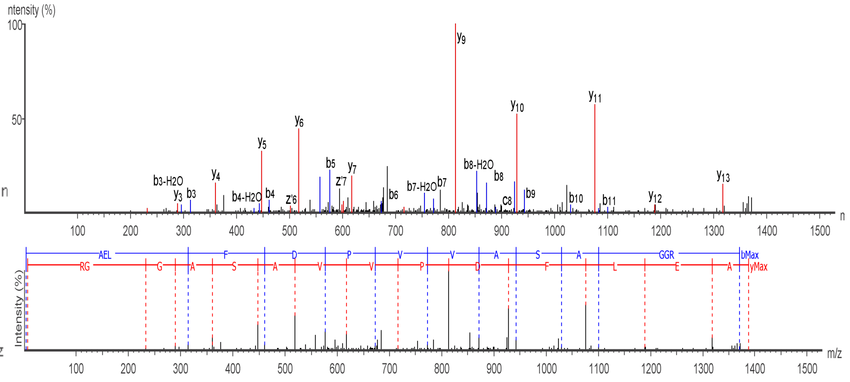
pplementary Figures

Supplementary Figure 1. Scheme for PCR amplification of target NRPS/PKS genes using reverse translated peptide sequences. Top panel shows a typical domain organization for an NRPS protein, where a condensation domain (C), an adenylation domain (A) and a thioation domain (T) are connected in tandem. The conserved regions of adenylation domains, A3 and A7, are also shown. A peptide matching to the NRPS is mapped to the domain context, and reverse translated into nuleotide sequences (peptide F and peptide R). PCR reactions can be performed using A3F/A7F as forward primer and peptide R as reverse primer (solid arrows), or on the other direction, using peptide F as forward primer and A3R/A7R from the next module as reverse primer (dashed arrows). The bottom panel shows a typical domain organization for a PKS protein, containing a ketosynthase domain (KS), a methyl-malonyl-CoA transferase (MAT) and an acyl-carrier protein (ACP), along with the conserved region K1 for KS and M6 for MAT. PCR reactions can be performed between K1F and peptide R as well as peptide F and M6R.

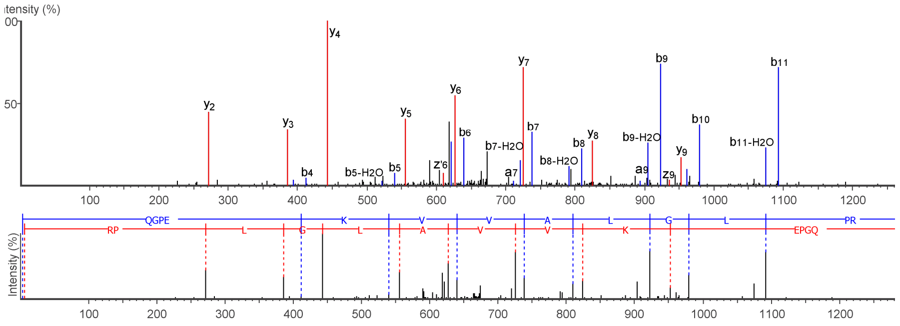
(a) *de novo* peptide identifications from strain F-6133



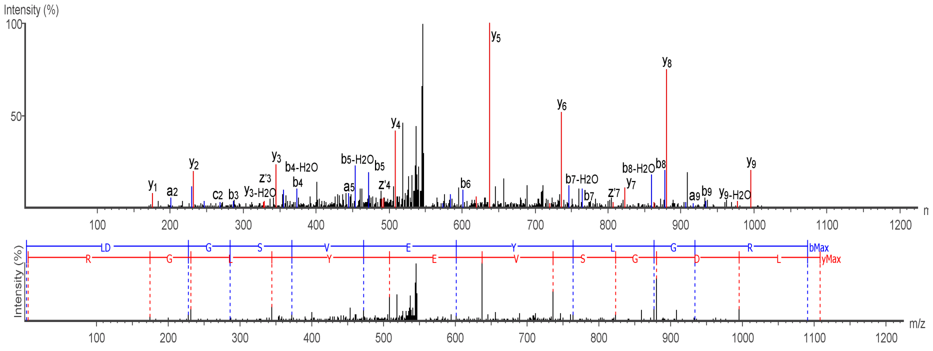
RFDLEAPPLLR



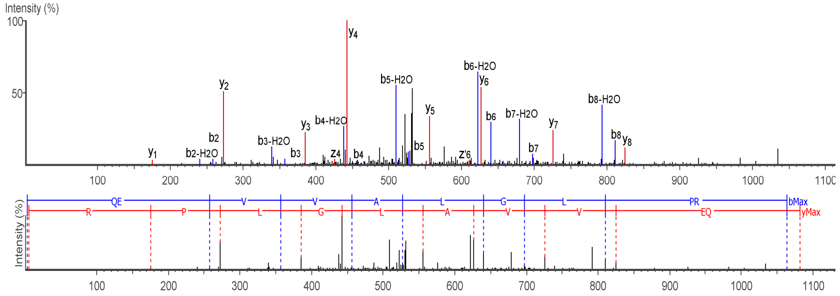
AELFDPVVASAG[GR]



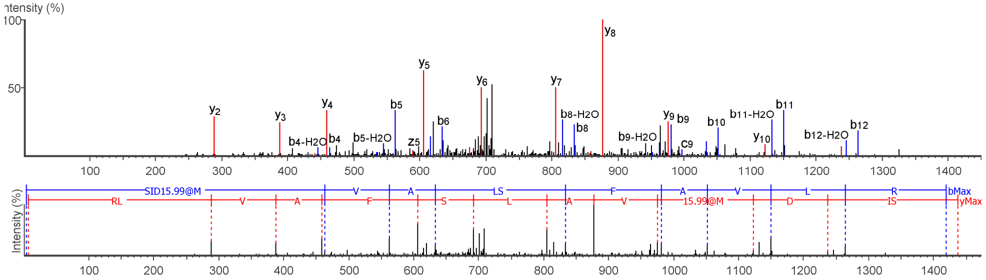
[QGPE]KVVALGL[PR]



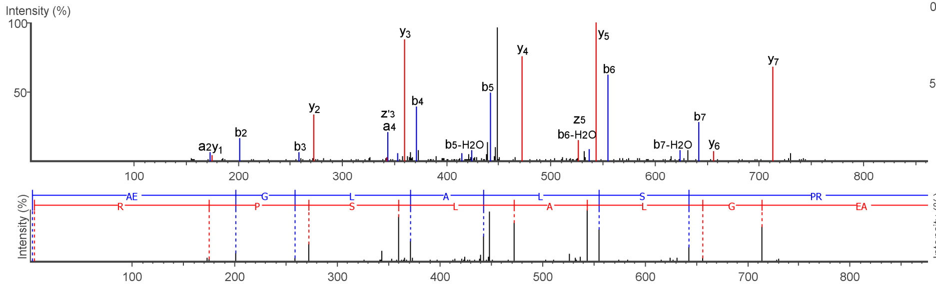
LDGSVEYLGR



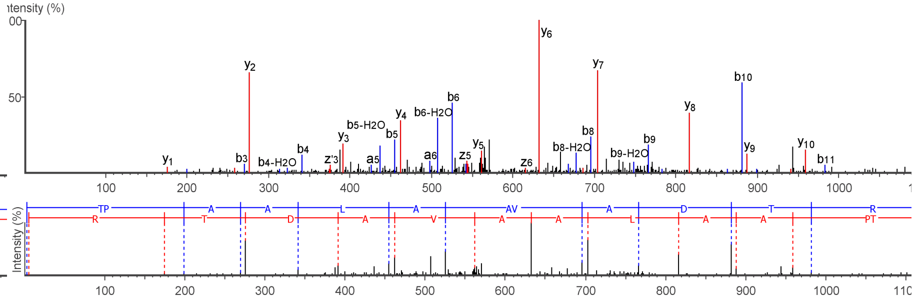
[QE]VVALGLPR



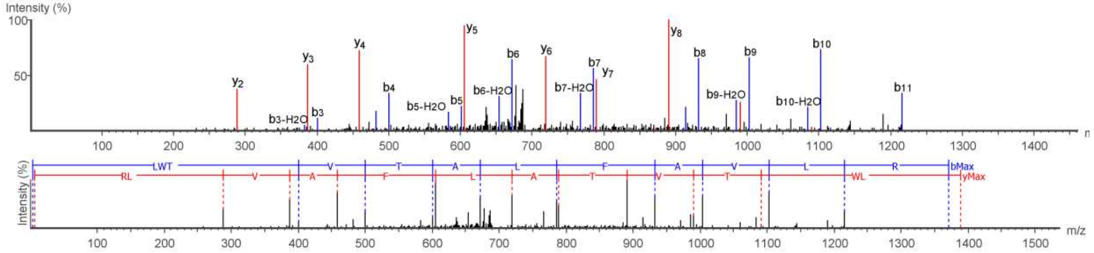
[SI]DmVALSFAVLR



[AE]GLALSPR

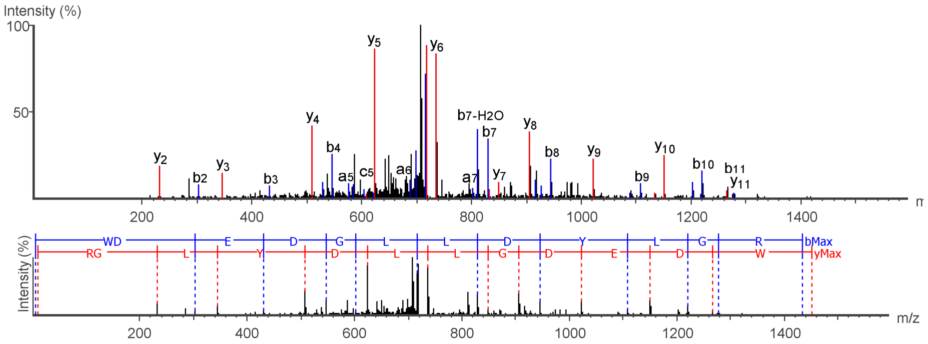


[TP]AALAAVADTR

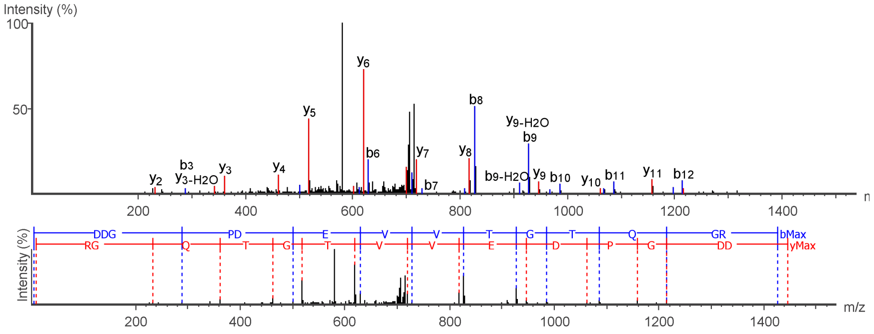


[LW]TVTALFAVLR

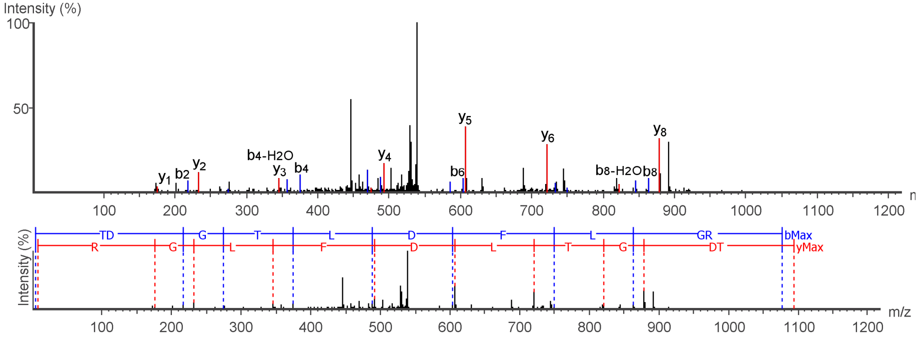
(b) *de novo* peptide identifications from strain F-6556



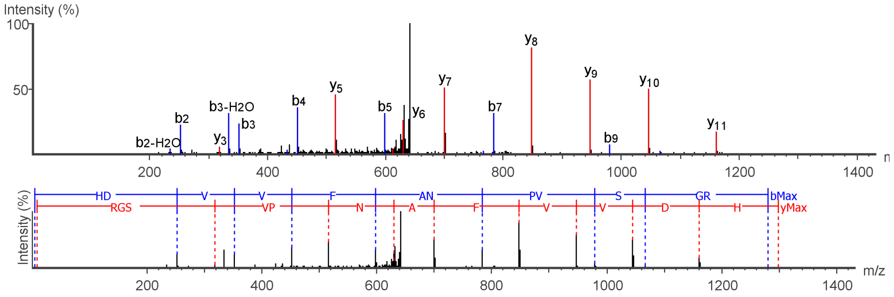
WDEDGLLDYLGR



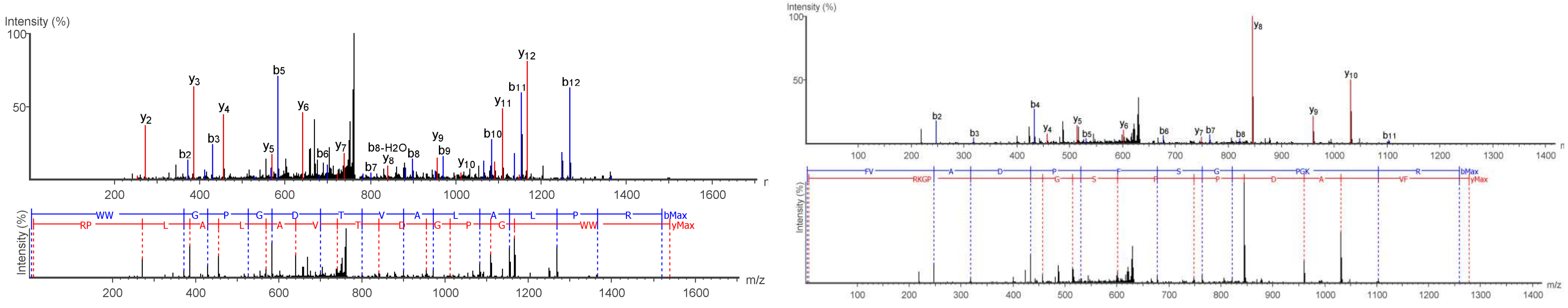
DDGPDEVVTGTQGR



[TD]GTLDFLGR



HDVVFGQPVSGR



[WW]GPGDTVALALPR

[FV]ADPFSG[GPK]R

Supplementary Figure 2. MS/MS spectrum of peptide identifications by PEAKS Studio are shown for strains **(a)** F-6133 and **(b)** F-6556. For each peptide identified by *de novo* sequencing, an original MS/MS spectrum is shown on the top panel, with *b* and *y* ions labeled, as well as an alignment to the proposed peptide sequence shown at the bottom panel. A '[]' indicates there is no fragment ion observed to clarify the sequence in the bracket. Other information of the identified peptides is shown in **Supplementary Table 2.**