1   **Isolation and characterization of Streptomyces bacteriophages and Streptomyces strains**

2   **encoding biosynthetic arsenals**

3

4   Streptomyces strains and phages for antibiotic discovery

5

6   Elizabeth T. Montaño, Jason F. Nideffer, Lauren Brumage, Marcella Erb, Alan I. Derman, John

7   Paul Davis, Elena Estrada, Sharon Fu, Danielle Le, Aishwarya Vuppala, Cassidy Tran, Elaine

8   Luterstein, Shivani Lakkaraju, Sriya Panchagnula, Caroline Ren, Jennifer Doan, Sharon Tran,

9   Jamielyn Soriano, Yuya Fujita, Pranathi Gutala, Quinn Fujii, Minda Lee, Anthony Bui, Carleen

10   Villarreal, Samuel R. Shing, Sean Kim, Danielle Freeman, Vipula Racha, Alicia Ho, Prianka Kumar,

11   Kian Falah, Thomas Dawson, Eray Enustun, Amy Prichard, Ana Gomez, Kanika Khanna, Shelly

12   Trigg, Kit Pogliano, Joe Pogliano*

13

14   Division of Biological Sciences, University of California, San Diego, La Jolla, California, USA.

15   * Corresponding author: jpogliano@ucsd.edu

Abstract

The threat to public health posed by drug-resistant bacteria is rapidly increasing, as some of healthcare's most potent antibiotics are becoming obsolete. Approximately two-thirds of the world's antibiotics are derived from natural products produced by Streptomyces encoded biosynthetic gene clusters. Thus, in order to identify novel gene clusters, we sequenced the genomes of four bioactive Streptomyces strains isolated from the soil in San Diego County and used Bacterial Cytological Profiling adapted for agar plate culturing in order to examine the mechanisms of bacterial inhibition exhibited by these strains. In the four strains, we identified 101 biosynthetic gene clusters. Some of these clusters were predicted to produce previously studied antibiotics; however, the known mechanisms of these molecules could not fully account for the antibacterial activity exhibited by the strains, suggesting that novel clusters might encode antibiotics. When assessed for their ability to inhibit the growth of clinically isolated pathogens, three Streptomyces strains demonstrated activity against methicillin-resistant *Staphylococcus aureus*. Additionally, due to the utility of bacteriophages for genetically manipulating bacterial strains via transduction, we also isolated four new phages (BartholomewSD, IceWarrior, Shawty, and TrvxScott) against *S. platensis*. A genomic analysis of our phages revealed nearly 200 uncharacterized proteins, including a new site-specific serine integrase that could prove to be a useful genetic tool. Sequence analysis of the Streptomyces strains identified CRISPR-Cas systems and specific spacer sequences that allowed us to predict phage host ranges. Ultimately, this study identified Streptomyces strains with the potential to produce novel chemical matter as well as integrase-encoding phages that could potentially be used to manipulate these strains.

Introduction

Antibiotic discovery is an international priority requiring immediate action.[1] The increasing prevalence of multi-drug resistant (MDR) bacterial pathogens has resulted in an increased use of last-resort antibiotics.[1-3] Microbes that produce natural products are the most prolific source of clinically approved antibiotics.[4] In particular, soil dwelling Actinobacteria, notably Streptomyces, account for two-thirds of the antibiotics currently on the market.[5-7] Despite intensive studies, however, the full potential of microbes to produce natural products has not been realized.[8] Genome mining studies have shown that microbes encode many biosynthetic gene clusters (BGCs) that have not yet been characterized.[8] It is widely assumed that many of these clusters produce novel natural products and that further characterization of Streptomyces bacteria increases the probability of identifying molecules with unique chemical structures and new mechanisms of action.[9]

In addition to identifying Streptomyces strains containing potentially novel BGCs, it is necessary to improve on the conventional approaches used in natural product antibiotic discovery. One of the major stumbling blocks in natural product discovery is dereplication, since the isolation of bioactive molecules often yields antibiotics that have previously been discovered.[10] We recently developed Bacterial Cytological Profiling (BCP) as a new whole-cell screening technique that can be used to rapidly identify the mechanism of action (MOA) of antibiotics.[11-16] BCP can accurately identify the pathway inhibited by antibacterial compounds present in unfractionated crude organic extracts and can be used to guide the purification of molecules with specific bioactivities.[11, 15] BCP can also be used to screen bacterial strains directly on petri plates in order to identify and prioritize those strains that produce molecules with desired MOA.[15] In effect, BCP helps with the problem of dereplication by allowing for the determination of the MOA of

62   antibiotics synthesized by a particular Streptomyces strain before labor-intensive antibiotic

63   purification efforts are performed.

64

65   Since many BGCs are not expressed under laboratory conditions, genetic methods are often

66   used to augment their expression and facilitate the identification and purification of their

67   products.[17] Sometimes, increased expression can be achieved using techniques such as

68   CRISPR/Cas or plasmid cloning and overexpression.[17] However, there is still an occasional need

69   to move large chromosomal regions from one strain to another via transduction in order to

70   engineer strains optimally suited for antibiotic production. Transduction requires a phage

71   capable of infecting the strain(s) of interest. Moreover, because phages generally display narrow

72   host ranges[18] and relatively few Streptomyces phages have been isolated[19] compared to the

73   large number of studied Streptomyces bacteria,[20] phages aptly suited for genetic manipulations

74   are not available for the majority of antibiotic producing Streptomyces strains isolated. In

75   addition, phage derived enzymes such as recombinases and integrases can also be used to

76   engineer new strains.[21-25]   Thus, studying the phages that infect antibiotic-producing

77   Streptomyces strains could not only yield new transducing phages but potentially also new

78   genetic tools for strain engineering.

79

80   Here we describe the isolation and characterization of Streptomyces strains and phages. We

81   used a combination of bioinformatics and BCP to characterize the antibiotic biosynthetic

82   potential of four Streptomyces strains that displayed an ability to inhibit Gram-negative and

83   Gram-positive growth. Additionally, we isolated four new phages and assessed their abilities to

84   infect our Streptomyces strains, which contained many CRISPRs. The proteins encoded by the

85   phages were subjected to bioinformatic analyses in order to identify putative integrases that

86   might be used for genetic manipulations. This work highlights a novel set of gene clusters and

87 Streptomyces sp. phages that serve as a starting point for the isolation of potentially novel

88 natural products.

89

90 **Results and Discussion**

91 *Isolation and Antibacterial Activities of Streptomyces sp.*

92 To identify Streptomyces strains containing potentially novel BGCs, we collected 28 unique soil

93 samples from sites across San Diego County. From these samples, we isolated a total of eight

94 bacterial strains that were determined, based on 16s RNA sequencing, to be members of the

95 Streptomyces genus (Fig 1, Table 1). These strains (designated JS, DF, QF2, EDE, SK, AH, ELW,

96 and SFW) and two known species (*Streptomyces coelicolor* A3(2) and *Streptomyces platensis*

97 AB045882) were screened using the cross-streak method for their ability to inhibit the growth of

98 wild type *E. coli* MC4100, an efflux defective mutant *E. coli* JP313 Δ*tolC,* and *B. subtilis* PY79.

99 Since the production of bioactive secondary metabolites is highly dependent on growth

100 conditions,[26] this screen was conducted on actinomycete isolation agar (AIA) as well as LB

101 agar. Each of the 10 strains proved capable of inhibiting the growth of *E. coli* and/or *B. subtilis*

102 on at least one of the tested media (Fig 2), suggesting that these strains likely produce

103 antibiotics. As expected, however, the production of antibiotics often depended upon whether

104 the strain was grown on AIA or LB agar. For example, strain ELW was incapable of inhibiting the

105 growth of Gram-negative and Gram-positive bacteria when grown on AIA. However, when

106 grown on LB agar, strain ELW inhibited the growth of both Gram-negative and Gram-positive

107 bacteria. Conversely, strains JS and QF2 exhibited growth inhibition regardless of the media on

108 which they were grown.

109

110 **Fig 1. Phylogenetic tree of Streptomyces bacteria isolated from soil samples.** This tree was
111 constructed using PCR-amplified 16S rRNA sequences and was generated from a multiple
112 sequence alignment using a neighbor-joining algorithm. The scale bar (0.03) indicates
113 substitutions per nucleotide.
114

115 **Table 1. Top NCBI BlastN hits of the 16S rRNA gene sequences.**

Sample ID	NCBI BlastN Top 16S ribosomal RNA Hit Description	Max Score	Total Score	Query Cover	E value	Percent Identity	Accession No.
S. platensis JCM 4662	Streptomyces platensis strain JCM 4662	2748	2748	100%	0	100%	NR_024761.1
S. coelicolor A3(2)	Streptomyces coelescens strain AS 4.1594	2793	2793	98%	0	99.93%	NR_027222.1
JS	Streptomyces rochei strain NRRL B-1559	2741	2741	91%	0	99.93%	NR_116078.1
DF	Streptomyces fulvissimus strain DSM 40593	2691	2691	100%	0	99.93%	NR_103947.1
QF2	Streptomyces californicus strain NBRC 12750	2699	2699	100%	0	100%	NR_112257.1
EDE	Streptomyces pratensis strain ch24	1238	1238	99%	0	98.99%	NR_125616.1
SK	Streptomyces californicus strain NBRC 12750	1242	1242	100%	0	100%	NR_112257.1
AH	Streptomyces pratensis strain ch24	1194	1194	100%	0	100%	NR_125616.1
ELW	Streptomyces atratus strain NRRL B-16927	1138	1138	100%	0	99.84%	NR_043490.1
SFW	Streptomyces caviscabies strain ATCC 51928	2767	2767	100%	0	99.87%	NR_114493.1

116

117

118 **Fig 2. Inhibition of bacterial growth by Streptomyces isolates.** The cross-streak method was
119 used to measure the zone of inhibition among ten Streptomyces strains against two Gram-
120 negative *E. coli* strains (MC4100 WT, JP313 Δ*tolC*), and Gram-positive *B. subtilis* PY79 on (**a**)
121 LB and (**b**) actinomycete isolation agar (AIA). Error bars represent the standard deviation of
122 three independent trials.
123

124 *Mechanistic Analysis of Natural Products Produced by Four Streptomyces Isolates*

125 Strains QF2, JS, SFW and DF all inhibited the growth of *E. coli* Δ*tolC* when grown on either AIA

126 or LB agar, but in each case, the mechanism underlying inhibition was unknown. Thus, we

127 utilized BCP to examine the mechanism of growth inhibition exhibited by the antibacterial natural

128 products synthesized by these four Streptomyces isolates. Each of the four strains was grown

129 on three different media (LB, AIA, or ISP2) for 5 days to allow for the synthesis and excretion of

130 natural products into the surrounding agar. We then added exponentially growing *E. coli* cells

131 adjacent to the Streptomyces lawn. After two hours of incubation at 30°C, the *E. coli* cells were

132 stained with fluorescent dyes and imaged with high resolution fluorescence microscopy. *E. coli*

133 cells incubated adjacent to each of the four Streptomyces isolates displayed characteristic

134 cytological profiles that, in some cases, allowed for the classification of these strains according

135 to the MOA of the natural products they produced (Fig 3).

136

Fig 3. BCP phenotypes of *E. coli* JP313 Δ*tolC* exposed to natural products produced by four Streptomyces soil isolates grown on different solid media. Also displayed, are *E. coli* JP313 Δ*tolC* untreated controls grown on the tested media (LB agar, AIA, and ISP2 agar). White arrows indicate cells with three chromosomes. BCP images were collected after staining the cells with FM4-64 (red), DAPI (blue), and SYTOX-green (green). The scale bar represents one micron.

142

When grown on either LB or ISP2, strain QF2 synthesized an antibiotic that caused the DNA of affected *E. coli* cells to assume a toroidal conformation (Fig 3). This phenotype is characteristic of bacteria treated with protein synthesis inhibitors such as chloramphenicol,[11, 27] and thus, we concluded that strain QF2 is capable of synthesizing a translation-inhibiting natural product. QF2 also produced a membrane-active secondary metabolite, evidenced by visible membrane abnormalities as well as Sytox Green permeability under all tested nutrient conditions (Fig 3). Strain JS appeared to induce similar phenotypes in *E. coli*, though under different growth conditions; protein synthesis inhibition phenotypes were observed on AIA and ISP2 but not on LB. And, as was the case with strain QF2, Sytox Green permeability was observed in some cells regardless of the medium on which strain JS was grown.

Strain SFW induced distinct phenotypes in *E. coli* cells under each of the three nutrient conditions (Fig 3). On LB, a significant number of *E. coli* cells grown in the presence of strain SFW appeared to contain three chromosomes (white arrows), a phenotype that was not present in the untreated control cells. When strain SFW was grown on AIA, the *E. coli* cells became bent and lost their characteristic rod shape. Finally, strain SFW grown on ISP2 induced substantial swelling in *E. coli* cells that ultimately led to lysis. Notably, *E. coli* cells grown in the presence of strain DF exhibited nearly these same phenotypes under identical growth conditions, suggesting that these two strains produce compounds targeting similar pathways.

162

163 *Genomic Analysis of Four Streptomyces Isolates*

164 In order to better understand how strains QF2, JS, SFW and DF inhibited bacterial growth, we

165 sequenced their genomes and aligned them to the most similar genomes in the NCBI database

166 (Fig 4A). Sequence reads for strain DF were assembled into a single contig that was most similar

167 to the genome of *S. fulvissimus* DSM 40593. Sequencing of strains QF2, JS, and SFW yielded

168 multiple contigs that were aligned to the genomes of *S. globisporus* C-1027, *S. parvulus* 12434,

169 *S. pratensis* ATCC 33331, respectively.

170

171 **Fig 4. Genome characteristics of Streptomyces strains DF, SFW, QF2, and JS isolated from soil**
172 **samples. (a)** Circularilzed representations of the linear genomes of the four bacterial isolates
173 represented by the assembled contigs obtained from genome sequencing. **(b)** Genomes oriented
174 according to their threonine operons (dark blue). Predicted biosynthetic gene clusters (light blue),
175 loci of Cas-associated protein-coding genes (green), and CRISPR arrays (purple) are shown. **(c)**
176 A Venn diagram displaying the numbers of BGCs that are shared by and unique to the genomes
177 of our isolates. Five clusters of particular importance are explicitly named.
178

179 In order to identify predicted gene clusters associated with secondary metabolism, the

180 assembled genome sequences for strains QF2, JS, SFW and DF were annotated using RASTtk[28]

181 and submitted to AntiSMASH 5.0[29] (Fig 4B). Each strain encoded between 18 and 34 BGCs,

182 some of which were present in multiple strains (Fig 4C). Additionally, some of the encoded

183 clusters closely resembled known BGCs in the MIBig repository.[30] For example, of the 23

184 putative BGCs identified in the genome of strain QF2 (Table 2), one of them (cluster 21) was

185 similar to the viomycin BGC (Fig 5A). Viomycin inhibits protein synthesis by stabilizing tRNAs in

186 the A site of the bacterial ribosome, inhibiting translocation.[31] According to AntiSMASH, 66% of

187 the genes within the viomycin BGC were similar to genes within cluster 21. However, a global

188 pairwise alignment of cluster 21 and the viomycin BGC revealed that the nucleotide sequence

189 of cluster 21 is actually 98.5% identical over 32.5kb of the 36kb viomycin BGC (Fig 5A). This

190 suggests that a viomycin related molecule is synthesized by strain QF2 and may account for

191   strain's ability to inhibit protein synthesis in *E. coli* (Fig 3). While some of the other clusters in the

192   QF2 genome (Table 2: clusters 2, 4, 7, 8, 9, 12, 13, 22) have significant similarity to known BGCs,

193   no other clusters appear to produce known antibiotics.

194

195   **Table 2. BGCs encoded within the draft genome sequence of strain QF2.**

Strain - QF2

Cluster	Type	Most Similar MiBIG Cluster and Predicted Percent Similarity	Antibacterial Activity	MIBig BGC-ID	Minimum	Maximum	Length (nt.)
1	butyrolactone	Coelimycin P1 T1PKS (8%)		BGC0000038	28795	39496	10702
2	terpene	Geosmin Terpene (100%)		BGC0001181	60346	82527	22182
3	NRPS	Griseobactin NRPS (35%)		BGC0000368	101394	123548	22155
4	NRPS	Coelichelin NRPS (72%)		BGC0000325	123549	145157	21609
5	T3PKS	Herboxidiene T1PKS, T3PKS (6%)		BGC0001065	184513	202853	18341
6	terpene	Isorenieratene Terpene (57%)		BGC0000664	681001	697369	16369
7	ectoine	Ectoine Other (75%)		BGC0000853	1125565	1134445	8881
8	T2PKS	Griseorhodin T2PKS (69%)		BGC0000230	1908342	1950906	42565
9	siderophore	Desferrioxamine B Siderophore (80%)		BGC0000941	2278919	2290697	11779
10	LAP,thiopeptide	-	-	-	2690041	2722548	32508
11	ectoine,butyrolactone	Showdomycin Other (47%)	Nucleic Acid and Protein Synthesis	BGC0001778	3344033	3359401	15369
12	melanin	Melanin Other (100%)		BGC0000911	4777563	4787988	10426
13	lanthipeptide	AmfS Lanthipeptide (100%)		BGC0000496	5105643	5127766	22124
14	terpene	-	-	-	5476141	5497007	20867
15	siderophore	Ficellomycin NRPS (3%)	DNA Replication	BGC0001593	5881146	5896034	14889
16	NRPS	Vioprolide A NRPS (25%)		BGC0001822	6093895	6137611	43717
17	bacteriocin	-	-	-	6194201	6205608	11408
18	NRPS-like	-	-	-	6430967	6442494	11528
19	NRPS-like,ladderane,arylpolyene	Skyllamycin NRPS (14%)	Unknown MOA	BGC0000429	6608620	6645073	36454
20	terpene	Hopene Terpene (46%)		BGC0000663	6755130	6764031	8902
21	NRPS,T1PKS	Viomycin NRPS (66%)	Protein Synthesis	BGC0000458	6806327	6870853	64527
22	T3PKS	Alkylresorcinol T3PKS (66%)	Unknown MOA	BGC0000282	7058484	7080991	22508
23	lassopeptide	-	-	-	7260793	7282889	22097

196
197   The most similar BGCs in the MiBIG database are listed, as well as the percentage of genes in
198   each MiBIG known cluster that have similarity to genes in the corresponding QF2 cluster. In
199   cases where the most similar known BGC produces an antibiotic, the MOA was listed
200   (Showdomycin,[32] Ficellomycin,[33] Skyllamycin,[34] Viomycin,[31] Alkylresorcinol[35]).
201

202   **Fig 5. Comparison of BGCs encoded in the genomes of bacterial soil isolates and the predicted
203   most similar, previously characterized BGC with an antibacterial product. (a)** Strain QF2, BGC
204   21, compared to the BGC previously described to produce the antibiotic viomycin (NCBI Acc
205   No. AY263398.1), encoded in the WGS of *S. vinaceus* ATCC 11861. **(b)** Strain JS, BGC 7,
206   compared to the BGC previously described to produce the antibacterial sesquiterpene
207   Albaflavenone (NCBI Acc No. AL645882.2), encoded in the WGS of *S. coelicolor* A3(2). **(c)** Strain
208   SFW, BGC 1, compared to the BGC previously described to produce the antibacterial beta-
209   lactam Carbapenem MM 4550 (NCBI Acc No. KF042303.1), encoded in the WGS of *S.
210   argenteolus* ATCC 11009. **(d)** Strain DF, BGC 27, compared to the BGC previously described to
211   produce the ammonium ionophore antibiotic Nonactin (NCBI Acc No. AF074603.2), encoded in
212   the WGS of *S. griseus subsp. griseus* ETH A7796. Cluster comparisons were constructed in
213   Easyfig. Regions of nucleotide homology are indicated on a gray scale and genes are colored
214   according to the putative function of the corresponding protein product.
215

216   Strain JS contained 18 putative BGCs, six of which shared significant similarity (>60% of genes

217   in common) with a known cluster (Table 3). Of these six, however, only cluster 7 was predicted

218 to produce an antibiotic. All of the genes constituting a known terpene cluster that produces

219 albaflavenone were present in cluster 7 (Fig 5B). Albalfavenone is capable of inhibiting the growth

220 of *B. subtilis* by an unknown MOA[36] and has previously been isolated from *S. coelicolor* A3(2),[37]

221 a close relative of strain JS. Since the MOA of albaflavenone is unknown, it's not clear whether

222 the products of cluster 7 or of a different cluster are responsible for the inhibition of protein

223 synthesis and/or the membrane permeability observed in *E. coli* (Fig 3).

224

225 **Table 3. BGCs encoded within the draft genome sequence of strain JS.**

Strain - JS

Cluster	Type	Most Similar MiBIG Cluster and Predicted Percent Similarity	Antibacterial Activity	MIBig BGC-ID	Minimum	Maximum	Length (nt.)
1	T3PKS	Herboxidiene T1PKS, T3PKS (7%)		BGC0001065	214109	229085	14977
2	ectoine	Ectoine Other (100%)		BGC0000853	672467	682865	10399
3	melanin	Melanin Other (60%)		BGC0000911	1414668	1425276	10609
4	siderophore	Desferrioxamine B Siderophore (66%)		BGC0000941	1507497	1519344	11848
5	furan	Methylenomycin Other (9%)	Potentially Inhibits Cell Wall Biosynthesis	BGC0000914	2667706	2688702	20997
6	NRPS	Ansamitocin P-3 T1PKS (7%)		BGC0001511	3021294	3076234	54941
7	terpene	Albaflavenone Terpene (100%)	Unknown MOA	BGC0000660	3743916	3764845	20930
8	T2PKS	Spore pigment T2PKS (66%)		BGC0000271	3800219	3857023	56805
9	siderophore	-	-	-	4274698	4286154	11457
10	bacteriocin	-	-	-	4471015	4482384	11370
11	terpene	-	-	-	4488880	4508472	19593
12	NRPS	Lipopeptide 8D1-1 & 8D1-2 NRPS (25%)	PMF Collapse	BGC0001370	4616241	4669719	53479
13	NRPS	Lipopeptide 8D1-1 & 8D1-2 NRPS (15%)	PMF Collapse	BGC0001370	4929292	4967094	37803
14	terpene	Hopene Terpene (76%)		BGC0000663	5042481	5069167	26687
15	terpene	Lysolipin T2PKS (4%)	Cell Wall Biosynthesis	BGC0000242	5088579	5103665	15087
16	T1PKS	Candicidin T1PKS (28%)		BGC0000034	5331251	5350185	18935
17	T2PKS,butyrolactone	Kinamycin T2PKS (25%)	DNA Synthesis	BGC0000236	5370217	5395391	25175
18	T1PKS	FR-008/Levorin A3 T1PKS (28%)		BGC0000061	5395392	5413565	18174

226
227 The most similar BGCs in the MiBIG database are listed, as well as the percentage of genes in
228 each MiBIG known cluster that have similarity to genes in the corresponding JS cluster. In cases
229 where the most similar known BGC produces an antibiotic, the MOA was listed
230 (Methylenomycin,[38] Albaflavenone,[36] Lipopeptide 8D1-1 & 8D1-2,[39] Lysolipin,[40] Kinamycin[41]).
231

232 Of the 26 putative BGCs that were identified in the genome of strain SFW, only one cluster shared

233 a high percentage of genes in common with a known antibiotic-producing cluster (Table

234 4). Cluster 1 shared similarity with 62% of the genes within a known BGC that produces

235 carbapenems (Fig 5C), a class of beta-lactam antibiotics that inhibit cell wall biogenesis.[42,]

236 [43] Additionally, cluster 4 contained a low percentage of genes in common with a BGC involved

237 in the synthesis of clavulanic acid, which inhibits beta-lactamase and consequently strengthens

238 the bactericidal activity of beta-lactams. Cluster 1 (and perhaps cluster 4) could, therefore,

239 contribute to the synthesis of bioactive molecules that account for the inhibition of *E. coli* cell

240 wall biogenesis on ISP2 media (Fig 3).

241

242 **Table 4. BGCs encoded within the draft genome sequence of strain SFW.**

Strain - SFW

Cluster	Type	Most Similar MiBIG Cluster and Predicted Percent Similarity	Antibacterial Activity	MIBig BGC-ID	Minimum	Maximum	Length (nt.)
1	NRPS, blactam	Carbapenem MM 4550 Other (62%)	Cell Wall Biosynthesis	BGC0000842	280243	420495	140253
2	NRPS	Coelichelin NRPS (90%)		BGC0000325	537068	587954	50887
3	terpene	Isorenieratene Terpene (28%)		BGC0000664	601924	615201	13278
4	blactam	Clavulanic acid Other (20%)	Beta-lactamase Inhibition	BGC0000845	839861	863248	23388
5	terpene	Hopene Terpene (69%)		BGC0000663	920123	946635	26513
6	T1PKS	Sceliphrolactam T1PKS (72%)		BGC0001770	1325302	1388343	63042
7	bacteriocin	-	-	-	1598693	1609196	10504
8	lanthipeptide	Kanamycin Saccharide (1%)	Protein Synthesis	BGC0000703	1734703	1760347	25645
9	NRPS	Lipopeptide 8D1-1 & 8D1-2 NRPS (6%)	PMF Collapse	BGC0001370	1773501	1830128	56628
10	siderophore	Ficellomycin NRPS (3%)	DNA replication	BGC0001593	2107372	2120491	13120
11	terpene	-	-	-	2186197	2205874	19678
12	butyrolactone	Lactonamycin T2PKS (3%)	Protein Synthesis	BGC0000238	4018281	4029096	10816
13	NRPS	Istamycin Saccharide (11%)	Protein Synthesis	BGC0000700	4251191	4307412	56222
14	siderophore	Desferrioxamine B Siderophore (83%)		BGC0000941	4924798	4936579	11782
15	lanthipeptide	-	-	-	5321361	5346350	24990
16	terpene	-	-	-	5588840	5608518	19679
17	ectoine	Ectoine Other (100%)		BGC0000853	6072388	6080990	8603
18	T2PKS, PKS-like	Cinerubin B T2PKS (25%)	DNA Intercalation	BGC0000212	6443938	6515214	71277
19	terpene	Steffimycin T2PKS-Saccharide (16%)		BGC0000273	6560271	6580717	20447
20	terpene, ectoine	Ectoine Other (100%)		BGC0000853	6860752	6881669	20918
21	bacteriocin	-	-	-	6909859	6920014	10156
22	T3PKS	Tetronasin T1PKS (11%)	PMF Collapse	BGC0000163	7071589	7112647	41059
23	melanin	Melanin Other (100%)		BGC0000911	7208921	7219385	10465
24	T2PKS, terpene	Spore pigment T2PKS (83%)		BGC0000271	7244899	7317424	72526
25	NRPS	Rimosamide NRPS (21%)		BGC0001760	7458615	7511513	52899
26	butyrolactone	-	-	-	7592249	7602533	10285

243
244 The most similar BGCs in the MiBIG database are listed, as well as the percentage of genes in
245 each MiBIG known cluster that have similarity to genes in the corresponding SFW cluster. In
246 cases where the most similar known BGC produces an antibiotic, the MOA was listed
247 (Carbapenem,[43] Clavulanic acid,[44] Kanamycin,[45] Lipopeptide 8D1-1 & 8D1-2,[39] Ficellomycin,[33]
248 Lactonamycin,[46, 47] Istamycin,[48-50] Cinerubin,[51] Tetronasin[52]).
249

250 Strain DF encoded 34 BGCs (Table 5). Despite this rich supply of BGCs, however, we were only

251 able to identify one cluster that likely participates in the synthesis of an antibiotic with a confirmed

252 MOA. According to AntiSMASH v5.0, cluster 27 shared 85% gene identity with a known BGC

253 that produces nonactin, a bioactive ionophore that disrupts membrane potential[53] (Fig 5D). The

254 known clusters could not fully account for the antibacterial activity exhibited by strain DF (Fig 3),

255 suggesting that antibiotics might be produced by novel clusters.

256

257 **Table 5. BGCs encoded within the closed genome sequence of strain DF.**

Strain - DF

Cluster	Type	Most Similar MiBIG Cluster and Predicted Percent Similarity	Antibacterial Activity	MIBig BGC-ID	Minimum	Maximum	Length (nt.)
1	ectoine	-	-	-	55194	65595	10402
2	butyrolactone	Gamma-butyrolactone (100%)		BGC0000038	156789	167693	10905
3	terpene	Geosmin Terpene (100%)		BGC0001181	196683	218891	22209
4	transAT-PKS, T1PKS, NRPS	Griseobactin NRPS (82%)		BGC0000368	227963	343592	115630
5	NRPS	Coelichelin NRPS (81%)		BGC0000325	363422	414309	50888
6	NRPS	Arsenopolyketides Other (41%)	Unknown MOA	BGC0001283	421648	466618	44971
7	T3PKS	Naringenin Terpene (100%)	Potentially Inhibits Fatty Acid Biosynthesis	BGC0001310	481558	522683	41126
8	T2PKS	Granaticin T2PKS (29%)		BGC0000227	874325	916838	42514
9	terpene	Steffimycin T2PKS-Saccharide (19%)		BGC0000273	1079402	1100700	21299
10	ectoine	Ectoine Other (100%)		BGC0000853	1578303	1588701	10399
11	NRPS	Laspartomycin NRPS (9%)	Cell Wall Biosynthesis	BGC0000379	2182777	2259643	76867
12	siderophore	Desferrioxamine B Siderophore (100%)		BGC0000941	2696558	2708336	11779
13	LAP, thiopeptide	-	-	-	3100604	3132306	31703
14	NRPS	Microsclerodermins NRPS, T1PKS (14%)		BGC0001019	3327941	3392550	64610
15	betalactone	Divergolide T1PKS (6%)		BGC0001119	3740002	3763908	23907
16	T2PKS	Rabelomycin T2PKS-Saccharide (25%)		BGC0000262	4318297	4360617	42321
17	lassopeptide	Keywimycin RiPP (100%)		BGC0001634	4419581	4442206	22626
18	T1PKS	Streptazone E T1PKS (91%)		BGC0001296	4973731	5038740	65010
19	lanthipeptide	AmfS Lanthipeptide (100%)		BGC0000496	5317536	5340199	22664
20	terpene	-	-	-	5677735	5698600	20866
21	siderophore	-	-	-	6150075	6164335	14261
22	butyrolactone	4-formylaminooxyvinylglycine Other (21%)		BGC0001488	6317636	6328541	10906
23	bacteriocin	-	-	-	6488518	6499849	11332
24	terpene	2-methylisoborneol Terpene (100%)		BGC0000658	6514280	6535368	21089
25	NRPS	Asukamycin T2PKS (12%)	Unknown MOA	BGC0000187	6618679	6679113	60435
26	arylpolyene	Tetrapetalone A-D PKS (26%)	Unknown MOA	BGC0001798	6726255	6767369	41115
27	NRPS	Nonactin T2PKS (85%)	Dissipates Transmembrane Electric Potential	BGC0000252	6774441	6840334	65894
28	terpene	Hopene Terpene (69%)		BGC0000663	7093211	7119781	26571
29	NRPS	SGR PTMs NRPS, T1PKS (100%)	Unknown MOA	BGC0001043	7204325	7247340	43016
30	bacteriocin	-	-	-	7260810	7271608	10799
31	melanin	Melanin Other (100%)		BGC0000911	7452699	7463166	10468
32	T3PKS	Herboxidiene T1PKS, T3PKS (7%)		BGC0001065	7495580	7536629	41050
33	terpene	Isorenieratene Terpene (100%)		BGC0000664	7626626	7652190	25565
34	thiopeptide, LAP, NRPS	Lactazole Thiopeptide (33%)		BGC0000606	7666072	7717669	51598

258
259   The most similar BGCs in the MiBIG database are listed, as well as the percentage of genes in
260   each MiBIG known cluster that have similarity to genes in the corresponding DF cluster. In cases
261   where the most similar known BGC produces an antibiotic, the MOA was listed
262   (Arsenopolyketides,[54] Naringenin,[55] Laspartomycin,[56] Asukamycin,[57] Tetrapetalone,[58] Nonactin,[53]
263   SGR PTMs NRPS[59]).
264

265   *Antimicrobial Activity of Four Streptomyces isolates Against Clinically Relevant Pathogens*

266   In order to assess the relevance of antibiotics produced by strains JS, DF, SFW, and QF2, we

267   screened their ability to inhibit the growth of three clinically isolated pathogens using the cross-

268   streak method (Table 6). Both strain QF2 and strain JS inhibited the growth of methicillin-

269   resistant *S. aureus* (MRSA) and efflux-deficient *P. aeruginosa* PA01. These strains did not,

270   however, inhibit the growth of the wild-type clinical isolates *P. aeruginosa* PA01 and *P.*

271   *aeruginosa* P4, which were resistant to the antibiotics produced by all four Streptomyces

272   isolates. Strain DF, though incapable of inhibiting the growth of *E. coli tolC⁺* (Fig 2), did inhibit

273   the growth of MRSA and efflux-deficient *P. aeruginosa* PA01. Strain SFW was the least capable

274   of inhibiting the growth of clinical pathogens, producing antibiotics only effective against *E. coli*

275   *tolC⁺* (Fig 2).

276

277   **Table 6. Inhibition of growth of clinically relevant pathogens by Streptomyces strains DF, SFW,**
278   **QF2, and JS.**

	Gram-Negative Bacteria			Gram-Positive Bacteria
	P. aeruginosa			*S. aureus*
	PAO1	P4	PAO1 Δefflux	MRSA USA300 TCH1516
DF	-	-	+	+
SFW	-	-	-	-
QF2	-	-	+	+
JS	-	-	+	+

279
280   Plus signs indicate growth inhibition, while minus signs indicate pathogen growth.

281

282   *Phage Isolation and Genome Sequencing*

283   With the goal of identifying genetic tools that could be used to augment expression of the BGCs

284   in our bacterial isolates, we decided to isolate bacteriophages using *S. platensis* as a host. This

285   species, in particular, was chosen as a host because it is relatively well-characterized, and *S.*

286   *platensis* phages capable of infecting our Streptomyces isolates could possibly be used to move

287   (via transduction) BGCs from our isolates into a more genetically manipulatable and familiar

288   background.[60, 61] Thus, to increase the probability that our *S. platensis* phages could be used

289   for this purpose, we performed the isolation using the same soil samples from which our

290   Streptomyces strains were obtained. Four *S. platensis* bacteriophages (BartholomewSD,

291   IceWarrior, Shawty, and TrvxScott) were successfully isolated. These phages were imaged using

292   negative-stain transmission electron microscopy (Fig 6A) and subsequently characterized as

293   members of the family Siphoviridae due to their long filamentous tails and icosahedral capsids.[62,]

294   [63] Genome sequencing revealed that BartholomewSD (52.1 kb) and TrvxScott (52.6 kb) are 89%

295   identical (Fig 6B) and belong to the BD2 subcluster of Streptomyces phages, which currently

296   contains 20 other members.[19] IceWarrior (55.5 kb) clustered in subcluster BI1 (24 members), and

297   Shawty (40.7 kb) clustered in BB1, a subcluster of 7 phages that includes notable members TG1

298   and phiC31 (Table 7).[19] The BLASTp-predicted functions of the gene products encoded by these

299   phages are shown in Table 8.

300

301   **Fig 6. Characterization of four Streptomyces phages isolated from soil samples. (a)** Electron
302   micrographs of the four phages (IceWarrior, TrvxScott, BartholomewSD, and Shawty). Lysate
303   samples were negatively stained and imaged with transmission electron microscopy (TEM). The
304   scale bar represents 100 nm. **(b)** A whole-genome sequence comparison of the four phages
305   generated by Phamerator (top to bottom: TrvxScott, BartholomewSD, Shawty, IceWarrior).
306

307   **Table 7. Summary of the NCBI WGS annotations of four phage isolates.**

Bacteriophage	Taxa	Genome (bp)	GC%	Genes	Cluster	Subcluster	Genbank Acc. No.
TrvxScott	unclass. Arequatrovirus	52600	67.8	81	BD	BD2	MH669016
IceWarrior	unclass. Rimavirus	55532	59.5	86	BI	BI1	MK433259
BartholomewSD	unclass. Arequatrovirus	52131	67.6	88	BD	BD2	MK460245
Shawty	unclass. Lomovskayavirus	40733	63.2	58	BB	BB1	MK433266

308

309

310

Table 8. Functions of the putative proteins encoded within the genomes of four phage isolates.

Streptomyces Phage TrvxScott taxon:2301575		Streptomyces phage Shawty taxon:2510521		Streptomyces phage IceWarrior taxon:2510515		Streptomyces phage BartholomewSD taxon:2510587	
CDS No.	Product	CDS No.	Product	CDS No.	Product	CDS No.	Product
1	hypothetical protein	1	terminase small subunit	1	hypothetical protein	1	hypothetical protein
2	HNH endonuclease	2	terminase large subunit	2	HNH endonuclease	2	hypothetical protein
3	thioredoxin	3	portal protein	3	hypothetical protein	3	HNH endonuclease
4	terminase small subunit	4	capsid maturation protease	4	hypothetical protein	4	tRNA-Phe
5	terminase large subunit	5	major capsid protein	5	endolysin	5	hypothetical protein
6	portal protein	6	head-to-tail adaptor	6	head-to-tail connector complex protein	6	thioredoxin
7	capsid maturation protease	7	hypothetical protein	7	hypothetical protein	7	hypothetical protein
8	scaffolding protein	8	major tail protein	8	terminase large subunit	8	terminase
9	major capsid protein	9	hypothetical protein	9	hypothetical protein	9	portal protein
10	head-to-tail connector complex protein	10	tail assembly chaperone	10	hypothetical protein	10	MuF-like minor capsid protein
11	head-to-tail connector complex protein	11	tail assembly chaperone	11	hypothetical protein	11	scaffolding protein
12	hypothetical protein	12	tape measure protein	12	portal protein	12	major capsid protein
13	head-to-tail connector complex protein	13	minor tail protein	13	hypothetical protein	13	head-to-tail adaptor
14	major tail protein	14	minor tail protein	14	capsid maturation protease	14	head-to-tail stopper
15	tail assembly chaperone	15	minor tail protein	15	hypothetical protein	15	hypothetical protein
16	tail assembly chaperone	16	minor tail protein	16	hypothetical protein	16	tail terminator
17	tape measure protein	17	hypothetical protein	17	major tail protein	17	major tail protein
18	minor tail protein	18	tail fiber	18	hypothetical protein	18	tail assembly chaperone
19	minor tail protein	19	lysin A	19	major tail protein	19	tail assembly chaperone
20	hypothetical protein	20	hypothetical protein	20	hypothetical protein	20	tape measure protein
21	hypothetical protein	21	deoxynucleoside monophosphate kinase	21	chitosanase	21	minor tail protein
22	hypothetical protein	22	immunity repressor	22	hypothetical protein	22	minor tail protein
23	minor tail protein	23	Cas4 family exonuclease	23	tape measure protein	23	hypothetical protein
24	hypothetical protein	24	hypothetical protein	24	minor tail protein	24	minor tail protein
25	lysin A	25	hypothetical protein	25	minor tail protein	25	hypothetical protein
26	hypothetical protein	26	hypothetical protein	26	hypothetical protein	26	hypothetical protein
27	hypothetical protein	27	hypothetical protein	27	hypothetical protein	27	hypothetical protein
28	hypothetical protein	28	hypothetical protein	28	hypothetical protein	28	lysin A
29	hypothetical protein	29	hypothetical protein	29	holin	29	hypothetical protein
30	exonuclease	30	HNH endonuclease	30	hypothetical protein	30	hypothetical protein
31	hypothetical protein	31	DNA primase	31	hypothetical protein	31	immunity repressor
32	hypothetical protein	32	restriction endonuclease	32	hypothetical protein	32	hypothetical protein
33	hypothetical protein	33	DNA polymerase I	33	hypothetical protein	33	Cas4 family exonuclease
34	deoxycytidylate deaminase	34	RNA polymerase sigma factor	34	hypothetical protein	34	hypothetical protein
35	DNA helicase	35	hypothetical protein	35	hypothetical protein	35	hypothetical protein
36	holliday junction resolvase	36	hypothetical protein	36	hypothetical protein	36	hypothetical protein
37	hypothetical protein	37	hypothetical protein	37	hypothetical protein	37	deoxycytidylate deaminase
38	DNA primase	38	hypothetical protein	38	hypothetical protein	38	DnaB-like helicase
39	DNA primase	39	hypothetical protein	39	hypothetical protein	39	holliday junction resolvase
40	hypothetical protein	40	ThyX-like thymidylate synthase	40	hypothetical protein	40	hypothetical protein
41	hypothetical protein	41	hypothetical protein	41	hypothetical protein	41	DNA primase
42	exonuclease	42	hypothetical protein	42	hypothetical protein	42	DNA primase
43	hypothetical protein	43	thioredoxin	43	hypothetical protein	43	hypothetical protein
44	HTH DNA binding protein	44	hypothetical protein	44	hypothetical protein	44	hypothetical protein
45	hypothetical protein	45	deoxycytidylate deaminase	45	hypothetical protein	45	hypothetical protein
46	ribonucleotide reductase	46	hypothetical protein	46	hypothetical protein	46	Mre11 family dsDNA break repair endo/exonuclease
47	DNA methylase	47	hypothetical protein	47	hypothetical protein	47	hypothetical protein
48	hypothetical protein	48	hypothetical protein	48	hypothetical protein	48	helix-turn-helix DNA binding protein
49	hypothetical protein	49	hypothetical protein	49	hypothetical protein	49	hypothetical protein
50	hypothetical protein	50	hypothetical protein	50	hypothetical protein	50	ribonucleotide reductase
51	HTH DNA binding protein	51	hypothetical protein	51	hypothetical protein	51	hypothetical protein
52	integrase	52	hypothetical protein	52	hypothetical protein	52	hypothetical protein
53	hypothetical protein	53	protein kinase	53	hypothetical protein	53	hypothetical protein
54	thymidylate synthase	54	integrase	54	hypothetical protein	54	hypothetical protein
55	hypothetical protein	55	tRNA-Asp	55	hypothetical protein	55	helix-turn-helix DNA binding protein
56	hypothetical protein	56	tRNA-Thr	56	hypothetical protein	56	integrase
57	hypothetical protein	57	hypothetical protein	57	hypothetical protein	57	hypothetical protein
58	hypothetical protein	58	HNH endonuclease	58	hypothetical protein	58	ThyX-like thymidylate synthase
59	hypothetical protein			59	DNA primase/polymerase	59	hypothetical protein
60	hypothetical protein			60	hypothetical protein	60	hypothetical protein
61	deoxynucleoside monophosphate kinase			61	hypothetical protein	61	hypothetical protein
62	hypothetical protein			62	hypothetical protein	62	hypothetical protein
63	hypothetical protein			63	hypothetical protein	63	hypothetical protein
64	hypothetical protein			64	hypothetical protein	64	hypothetical protein
65	hypothetical protein			65	hypothetical protein	65	deoxymononucleoside kinase
66	hypothetical protein			66	hypothetical protein	66	hypothetical protein
67	hypothetical protein			67	hypothetical protein	67	hypothetical protein
68	hypothetical protein			68	hypothetical protein	68	hypothetical protein
69	hypothetical protein			69	hypothetical protein	69	hypothetical protein
70	hypothetical protein			70	hypothetical protein	70	hypothetical protein
71	hypothetical protein			71	hypothetical protein	71	hypothetical protein
72	hypothetical protein			72	hypothetical protein	72	hypothetical protein
73	hypothetical protein			73	hypothetical protein	73	hypothetical protein
74	hypothetical protein			74	hypothetical protein	74	hypothetical protein
75	hypothetical protein			75	hypothetical protein	75	hypothetical protein
76	hypothetical protein			76	hypothetical protein	76	hypothetical protein
77	hypothetical protein			77	hypothetical protein	77	hypothetical protein
78	acetyltransferase			78	hypothetical protein	78	hypothetical protein
79	hypothetical protein			79	hypothetical protein	79	hypothetical protein
80	hypothetical protein			80	hypothetical protein	80	hypothetical protein
81	hypothetical protein			81	DNA helicase	81	hypothetical protein
				82	HNH endonuclease	82	hypothetical protein
				83	hydrolase	83	hypothetical protein
				84	DNA helicase	84	hypothetical protein
				85	helix-turn-helix DNA binding domain protein	85	hypothetical protein
				86	hypothetical protein	86	hypothetical protein
						87	hypothetical protein
						88	hypothetical protein

311

312

313 *Characterization of CRISPR Elements in the Genomes of Our Streptomyces Strains*

314 Prior to testing the ability of our phages to infect the Streptomyces isolates, we decided to

315 examine the strains for complete and functional CRISPR/Cas systems. Our reasoning for this

316 was two-fold. First, the presence of acquired spacers and their specific sequences would allow

317 us to make predictions about whether or not our phages can infect our antibiotic-producing

318 strains. And, second, it was conceivable that in doing so we might discover a novel CRISPR/Cas-

319 based system. Our bioinformatic analysis identified the presence of Cas enzymes and CRISPR

320 arrays within the genomes all four of our Streptomyces isolates, but the abundance of CRISPRs

321 in each strain varied greatly (Table 9). QF2 contained the largest number of predicted CRISPRs

322 – 39 in total, scattered around the chromosome, each containing between one and 25 spacers

323 (Fig 4B, purple; Table 10). Some predicted spacers within these arrays matched with 94-100%

324 identity to sequences within TrvxScott (7 spacers), BartholomewSD (4 spacers), Shawty (2

325 spacers), and IceWarrior (5 spacers) (Table 10, Table 11). Spacers targeted a variety of genes

326 including those encoding minor tail proteins, tape measure proteins, deoxycytidylate deaminase,

327 helix-turn-helix DNA binding proteins, endolysin, and capsid maturation protease (Fig 7). The

328 large number of putative spacers in the QF2 genome targeting TrvxScott, BartholomewSD,

329 Shawty, and IceWarrior suggests that strain QF2 has likely previously encountered and acquired

330 resistance to each of these phages. Moreover, strain QF2 was isolated from the same soil sample

331 as BartholomewSD, providing support for these findings. Strain QF2 also encoded seven

332 proteins of a Type IE CRISPR-Cas system.[64-68] The QF2 proteins were distantly related to the

333 enzymes of the canonical Cas3 system in *E. coli* (Fig 8), but the operon in strain QF2 lacked two

334 genes (Cas1 and Cas2) involved in spacer acquisition. This phenomenon, the absence of Cas1

335 and Cas2, has previously been reported as a common feature of Streptomycetaceae Type IE

336 systems.[69] The presence in the QF2 genome of a Cas3 system and spacers mapping to essential

337   proteins in each of the genomes of our phages suggests that the strain is likely resistant to all

338   four of our phages, and thus, transduction is unlikely with strain QF2.

339

340   **Table 9. General characteristics of predicted CRISPR-Cas systems within the genomes of strains**
341   **DF, SFW, QF2, and JS.**

Strains	CRISPR	Spacers	Repeats	Spacers with Blastn Hits to Host Range Phage	Cas Loci	Cas-Associated Genes
DF	12	14	26	2	3	9
SFW	11	23	34	2	3	17
QF2	39	168	207	14	5	22
JS	4	8	12	0	4	20

342
343   Included in this table is the number of spacers within the genome of each bacterial strain with
344   sequence similarity to regions within any of the four phage isolates (IceWarrior, TrvxScott,
345   BartholomewSD, or Shawty).

346 **Table 10. Characteristics of the 38 CRISPRs predicted in the draft genome sequence of strain**
347 **QF2.**

Strain QF2

CRISPRs	Min	Max	Length (nt.)	No. Repeats	No. Spacers	Spacer Blastn Hit to Host Range Phage
1	384313	384428	116	2	1	S1 [BartholomewSD]
2	437656	437755	100	2	1	
3	835820	835922	103	2	1	
4	1307659	1307752	94	2	1	
5	1344898	1345205	308	6	5	
6	1543053	1543144	92	2	1	
7	1616159	1616542	384	5	4	S13 [BartholomewSD, TrvxScott]
8	1618208	1618291	84	2	1	
9	1833873	1833977	105	2	1	
10	1861097	1861197	101	2	1	
11	2316101	2316187	87	2	1	
12	2704894	2704990	97	2	1	
13	3015981	3016376	396	7	6	S21 [IceWarrior]
14	3106815	3107262	448	8	7	S27 [BartholomewSD, TrvxScott]
15	3112452	3113433	982	17	16	S34 [IceWarrior], S41 [TrvxScott]
16	3138610	3140161	1,552	26	25	
17	3145439	3145830	392	7	6	
18	3444685	3444779	95	2	1	
19	3507440	3507598	159	3	2	
20	3791739	3792012	274	5	4	
21	3838730	3838804	75	2	1	
22	4257871	4258080	210	4	3	S89 [IceWarrior]
23	4327550	4328916	1,367	23	22	S105 [IceWarrior], S106 [Shawty, TrvxScott]
24	4333365	4334666	1,302	22	21	S118 [TrvxScott] , S131 [Shawty, TrvxScott]
25	4335904	4336481	578	10	9	S138 [BartholomewSD], S140 [TrvxScott]
26	4522773	4522879	107	2	1	
27	4528080	4528148	69	2	1	
28	4657265	4657374	110	2	1	
29	4754273	4754356	84	2	1	
30	4787509	4787642	134	3	2	
31	4987714	4987810	97	2	1	
32	5305650	5305745	96	2	1	
33	5400452	5400541	90	2	1	S151 [IceWarrior]
34	5417083	5417162	80	2	1	
35	5441923	5442032	110	2	1	
36	6552625	6552699	75	2	1	
37	6798173	6798309	137	3	2	
38	7177566	7177797	232	6	5	

348
349 Spacers with sequence similarity to any of the four phages in this study are listed next to their
350 corresponding CRISPR and are identified according to their position relative to all other spacers
351 within the QF2 genome.
352

353

354 **Table 11. Spacers within the genomes of strain DF, SFW, and QF2 that have sequence similarity**
355 **to at least one of the four phage isolates.**

Strain DF

CRISPR	Spacer	Blastn Hit to Host Range Phage	Score (bits) / E Val.	No. Identities (%ID)	Strand	Minimum	Maximum	Length (nt.)	Sequence
9	S11	BartholomewSD	30.2 (15) / 2.4	15/15 (100%)	Plus / Minus	4171019	4171050	32	TGCCCACCGGCCGAGCCGCCTTCCGCAGGCAG
10	S12	TrvxScott BartholomewSD	30.2 (15) / 4.9 28.2 (14) / 9.5	15/15 (100%) 14/14 (100%)	Plus / Plus Plus / Minus	4684554	4684602	49	GGTGTCCCCGCCGGTCGCGTGCATGTCCTTCGGCTTGAGCGGGCTGCCG

Strain SFW

CRISPR	Spacer	Blastn Hit to Host Range Phage	Score (bits) / E Val.	No. Identities (%ID)	Strand	Minimum	Maximum	Length (nt.)	Sequence
1	S1	Shawty	28.2 (14) / 5.9	14/14 (100%)	Plus / Plus	250625	250650	26	GAGTCACCAGCCGGGCGAAGGCACGC
5	S6	IceWarrior	32.2 (16) / 0.99	19/20 (95%)	Plus / Plus	4553831	4553872	42	CGGGCGTCGACGGTGACGAGCGTCGCGTTCGTACTTCTCCTTG

Strain QF2

CRISPR	Spacer	Blastn Hit to Host Range Phage	Score (bits) / E Val.	No. Identities (%ID)	Strand	Minimum	Maximum	Length (nt.)	Sequence
1	S1	BartholomewSD	32.2 (16) / 1.2	16/16 (100%)	Plus / Plus	384347	384394	48	GCGGACGGCGGCGCGGCCGGTACCCCCGGTGTCCACGACGGCGGCGCG
7	S13	BartholomewSD TrvxScott	32.2 (16) / 1.9 32.2 (16) / 1.9	16/16 (100%) 16/16 (100%)	Plus / Plus Plus / Plus	1616356	1616423	68	CGACCTGCGGTACCACTCGATCCGGGCGCGGTCCCATCTACAAGGGCACGGTCGTCCAGCGGACCGAG
13	S21	IceWarrior	30.2 (15) / 2.5	15/15 (100%)	Plus / Minus	3016077	3016109	33	CGCCGGAACCCTCAAGGAGGAGAACGGCGCGGG
14	S27	BartholomewSD TrvxScott	30.2 (15) / 3.1 30.2 (15) / 3.1	18/19 (94%) 18/19 (94%)	Plus / Minus Plus / Minus	3106902	3106938	37	AGGGCCTGGCCGTGCGGGGTGCGGGTGGAGTCGTGGT
15	S34	IceWarrior	30.2 (15) / 2.4	15/15 (100%)	Plus / Plus	3112541	3112572	32	ACAGCGACGTCGCCTACAACTACGCCGCCTGG
15	S41	TrvxScott	28.2 (14) / 9.5	14/14 (100%)	Plus / Plus	3112961	3112992	32	GGTGCTGAACCCGTCGGCGGCCGTGAACTTGT
22	S89	IceWarrior	30.2 (15) / 2.5	15/15 (100%)	Plus / Plus	4257958	4257990	33	CCGCGGGCGTCCTTCGCCGAGGAGACCCTGCCC
23	S105	IceWarrior	30.2 (15) / 2.5	15/15 (100%)	Plus / Minus	4328429	4328461	33	CATCAGCGTCTGAAGCAGCACGCCCATCGCCTT
23	S106	Shawty TrvxScott	28.2 (14) / 9.5 28.2 (14) / 9.5	14/14 (100%) 14/14 (100%)	Plus / Plus Plus / Minus	4328491	4328522	32	TGGATCGAGCCGGACGGGCACATCAGCGGCCC
24	S118	TrvxScott	28.2 (14) / 9.5	14/14 (100%)	Plus / Plus	4333692	4333723	32	GCCGCGTCCGGCTACGGCTACGGCTCCGCCCC
24	S131	Shawty TrvxScott	28.2 (14) / 9.5 28.2 (14) / 9.5	14/14 (100%) 14/14 (100%)	Plus / Plus Plus / Minus	4334483	4334514	32	AACGCCGTCCATGAGGCGCTGCGTTTGGCGTC
25	S138	BartholomewSD	28.2 (14) / 9.5	14/14 (100%)	Plus / Plus	4336177	4336208	32	AACGCGGCAGCGATGGCCCGTACGAGCGGCGG
25	S140	TrvxScott	28.2 (14) / 9.5	14/14 (100%)	Plus / Minus	4336299	4336330	32	ATCCTCGCCGTCCAGACCGCCTCGACGCAGAT
33	S151	IceWarrior	36.2 (18) / 0.063	18/18 (100%)	Plus / Minus	5400476	5400517	42	GTGGTGGCCTCGCCGACCAGTTGCTCGGACGCCTGGGCGGCC

356
357 The bold portion of each spacer shares high sequence similarity with a region in the genome of
358 the listed host range phage. In cases where a single spacer mapped to two phages, bold and
359 underlined are used to sequences distinguish the two.

360

361 **Fig 7. Genomic maps of phages showing regions containing sequence similarity to spacers**
362 **found within the CRISPRs of strains QF2, DF, and SFW. (a)** TrvxScott, **(b)** BartholomewSD, **(c)**
363 Shawty, and **(d)** IceWarrior. Key displays putative functions of CRISPR targeted genes.
364

365 **Fig 8. Class I, Type I-E CRISPR-Cas system encoded in the WGS of strain QF2. (top)** The Type
366 I-E CRISPR-Cas operon encoded by strain QG is located from 4,973,482 to 4,987,810 and
367 includes seven genes. The Type I-E cascade is followed by CRISPR 31, consisting of two repeats
368 and a single spacer. **(bottom)** The canonical Type I-E CRISPR-Cas system encoded in the
369 genome of *E. coli* K-12 MG1655 is located from 2,887,219 to 2,877,618 and includes eight
370 genes. The Type I-E cascade is followed by a CRISPR 31, consisting of five repeats and four
371 spacers.
372

373 Specific spacers mapping to some of our phages were also discovered within the genomes of

374 strains DF and SFW (but not JS). Strain DF contained two spacers that mapped to sequences

375 within the genome of BartholomewSD, and one of these spacers also shared sequence similarity

376 with a region in TrvxScott (Table 11, Table 12). Strain SFW contained two spacers – one that

377 shared sequence similarity with Shawty and another that mapped to a sequence in IceWarrior

378 (Table 11, Table 13). Both strains DF and SFW encoded proteins containing regions with

379   similarity to the RuvC and HNH endonuclease domains of known Cas enzymes. However, given

380   the limited similarity of these putative proteins to known Cas proteins, further study is necessary

381   to determine if they constitute novel Cas systems. If these systems are functional, we predict

382   that strain DF is resistant to infection by TrvxScott and BartholomewSD, and strain SFW,

383   resistant to Shawty and IceWarrior.

384

385   **Table 12. Characteristics of the 12 CRISPRs predicted in the complete genome sequence of**
386   **strain DF.**

Strain DF

CRISPRs	Min	Max	Length (nt.)	No. Repeats	No. Spacers	Spacer Blastn Hit to Host Range Phage
1	1491701	1491797	97	2	1	
2	1517080	1517203	124	2	1	
3	1669449	1669522	74	2	1	
4	1995371	1995463	93	2	1	
5	2013671	2013864	194	3	2	
6	2241854	2241958	105	2	1	
7	2443114	2443225	112	2	1	
8	3767692	3767796	105	2	1	
9	4170935	4171076	142	3	2	S11 [BartholomewSD]
10	4684531	4684625	95	2	1	S12 [BartholomewSD, TrvxScott]
11	4802187	4802274	88	2	1	
12	5682554	5682639	86	2	1	

387
388   Spacers with sequence similarity to any of the four phages in this study are listed next to their
389   corresponding CRISPR and are identified according to their position relative to all other spacers
390   within the DF genome.
391

392   **Table 13. Characteristics of the 11 CRISPRs predicted in the draft genome sequence of strain**
393   **SFW.**

Strain SFW

CRISPRs	Min	Max	Length (nt.)	No. Repeats	No. Spacers	Spacer Blastn Hit to Host Range Phage
1	250601	250674	74	2	1	S1 [Shawty]
2	452415	452685	271	3	2	
3	2779718	2779796	79	2	1	
4	2824197	2824275	79	2	1	
5	4553801	4554168	368	5	4	S6 [IceWarrior]
6	4731148	4731651	504	10	9	
7	5166339	5166424	86	2	1	
8	5317268	5317371	104	2	1	
9	5722922	5723016	95	2	1	
10	6532966	6533033	68	2	1	
11	7183989	7184080	92	2	1	

394
395   Spacers with sequence similarity to any of the four phages in this study are listed next to their
396   corresponding CRISPR and are identified according to their position relative to all other spacers
397   within the SFW genome.

398

399   A curious feature emerged from our analysis of the CRISPRs within the Streptomyces strains.

400   Among the 213 predicted spacers encoded by all four bacterial strains, 18 contained sequence

401   similarity (95-100% identity) with at least one of the four phages (Table 9). The lengths of the

402   matching sequences (100% identity) within bacterial spacers ranged from 14 to 18 nucleotides

403   and accounted for approximately half the length of a typical spacer. Additionally, a single spacer

404   occasionally appeared capable of targeting two distantly related phages. These spacers

405   contained sequences mapping to two distinct genes encoded by different viral genomes. For

406   example, spacer 106 in CRIPSR 23 of strain QF2 is 32 nucleotides in length, and it contains 14

407   bases that share 100% identity with a region in the TrvxScott tape measure gene. These 14

408   bases overlap (by 8 nucleotides) with another sequence that is 14 base pairs in length and shares

409   100% identity with a region within the Shawty genome (Table 11). If these spacers functionally

410   serve to resist infection, our analysis suggests that a single spacer may evolve to efficiently target

411   more than one phage, thus providing broad immunity.

412

413   *Susceptibility of Streptomyces strains to infection by S. platensis phages*

414   With the hope of identifying phages that might serve as tools for transduction, we assessed the

415   susceptibility of our Streptomyces isolates to infection by each of the four *S. platensis*

416   phages. As predicted, strain QF2, with its Type IE Cas system and many CRISPRs containing

417   spacers against our phages, could not be infected by any of our four phages. Strain DF

418   experienced inefficient infection by TrvxScott (~2.0 x 10^4-fold reduced plating efficiency

419   compared to *S. platensis*) and was completely resistant to infection by BartholomewSD. In

420   addition to these results, which were generally predicted by our CRISPR/Cas findings, we also

421   demonstrated strain DF's resistance to infection by Shawty and susceptibility to IceWarrior (~20-

422   fold reduced efficiency). Strain SFW was at least partially resistant to infection by all four phages:

423   Shawty (no infection), BartholomewSD (no infection), IceWarrior (~10^7-fold reduced efficiency),

424   and TrvxScott (~1.3×10^5-fold reduced efficiency). Finally, strain JS, despite having no spacers

425   specifically targeting our phages, was similarly immune to infection by Shawty and

426   BartholomewSD and partially resistant to infection by IceWarrior and TrvxScott (~10^6-fold and

427   ~10^5-fold reduced efficiency, respectively).   These data are consistent with our predictions

428   regarding the resistance of our Streptomyces isolates to infection by the phages against which

429   they carry spacers, though it is the case that the presence of a spacer did not always confer

430   complete immunity to the phage it targeted. In some cases, strains containing spacers could be

431   infected relatively inefficiently by the targeted phage. For example, strain DF encoded a single

432   spacer targeting TrvxScott but remained partially susceptible to infection. Strains DF, SFW, and

433   JS were all capable of being infected by TrvxScott and IceWarrior to some degree. Thus, it

434   remains possible that these two phages could be used for transducing BGCs into *S. platensis*.

435

436   *Identification of phage integrases*

437   In analyzing the proteins encoded within the genomes of our phages, we identified site-specific

438   serine recombinases encoded by BartholomewSD, TrvxScott, and Shawty. The integrases of

439   BartholomewSD and TrvxScott were nearly identical and shared similarity to integrases

440   belonging to a number of previously studied phages. The Shawty integrase shared protein

441   sequence similarity with the integrases of Streptomyces phages TG1 and phiC31 (71.3% and

442   51.9% respectively). The TG1 and phiC31 integrases are distinct recombinases that share

443   49.7% protein sequence identity and have been used extensively as tools for integrating genes

444   of interest into specific loci within the genomes of a wide variety of organisms, ranging from soil

445   microbes such as Streptomyces to multicellular animals such as Drosophila.[21-25] Thus, as a

446   newly discovered member of this serine integrase family, the Shawty integrase could also

447   potentially be used to move Streptomyces-encoded BGCs between strains in order to facilitate

448   augmented expression of bioactive natural products.

449

450   This work demonstrates a method for effectively studying newly isolated, antibiotic-producing

451   bacteria and phages that may infect them. In particular, we have highlighted how BCP can be

452   used to assess the novelty of BGCs encoded within Streptomyces strains, providing a solution

453   to the problem of dereplication. Moreover, our work illustrates how the isolation and genomic

454   analysis of phages that infect antibiotic-producing Streptomyces strains might yield new genetic

455   tools such as transducing phages or integrases, which can be used to augment the expression

456   of novel antibiotics. Ultimately, our identification of potentially novel BGCs and phage integrases

457   serves as a foundation for further studies that could lead to the discovery of new antibiotics.

458

459   **Materials and Methods**

460   *Soil sample collection and site description*

461   Undergraduate students enrolled in the Phage Hunters Advancing Genomics and Evolutionary

462   Science (PHAGE) class at UCSD collected soil samples for isolating bacteria and their associated

463   phages. Soil samples (approx. 30 ml) were collected around San Diego County (32.7157° N,

464   117.1611° W), California, USA.

465

466   *Isolation of Streptomyces*

467   Actinomycete isolation agar (AIA) plates (for one liter: sodium caseinate 2 g, L-Asparagine 0.1 g,

468   sodium propionate 4 g, dipotassium phosphate 0.5 g, magnesium sulphate 0.1 g, glycerol 5 mL,

469   rifampicin 50 μgmL^{-1}, nystatin 100 μgmL^{-1}, cycloheximide 100 μgmL^{-1}, agarose 15 g, pH 8.1)

470   were used to select for Actinobacteria from soil samples. One gram of soil was added to the

471   agar surface and streaked across the AIA plate and incubated for two days at 30°C. The plates

472   were investigated for individual colonies with morphologies indicative of Streptomyces

473   (vegetative hyphae, aerial mycelium), and those colonies were picked and purified at least four

474   times on AIA-CHX plates.

475

476   *Phage isolation and purification*

477   Bacteriophages were isolated from soil samples with host, *Streptomyces platensis* JCM 4664

478   substr. MJ1A1. An enrichment culture was prepared from 1 g soil and 2.5 ml of *S. platensis*

479   added to 15 ml of Luria-Bertani (LB) medium (for one liter: tryptone 10 g, yeast extract 5 g, NaCl

480   10 g, agar 15 g, pH 7.0), followed by a 2-day incubation at 30°C with shaking. Phage were

481   isolated from a 1.2 ml volume of enrichment culture that was centrifuged at maximum speed for

482   3 min, 1 ml of the resulting supernatant was filtered (0.22 μM filter), and 5 μl of the filtrate was

483   spotted and then streaked onto an LB plate containing 100 μg/ml of cycloheximide. *S. platensis*

484   (0.1 ml) was mixed with 4.5 ml of LB top agar 0.7%, poured over the streak plate and incubated

485   for two days at 30°C. Resultant plaques were re-streaked onto new LB plates containing 100

486   μg/ml of cycloheximide about 3-4 times for phage purification.

487

488   *Bacterial genomic DNA extraction and quantification for 16S rRNA PCR amplification and*

489   *sequencing*

490   An adaptation of the DNeasy® Blood & Tissue Kit (Qiagen) protocol was used for bacterial

491   genomic DNA extraction. Strains were cultured overnight at 30°C in 5 ml of LB broth while rolling.

492   Cells were pelleted (16,000 x g, 3 min) from 1 ml of culture, re-suspended in 180 μl of lysis buffer

493   (prepared in house), and incubated at 37°C for 45 min after which 25 μl of proteinase K (20

494   mg/mL) and 200 μl Buffer AL (Qiagen) was added. The samples were vortexed at maximum

495   speed for 20 sec, incubated at 56°C for 30 min, and 200 μl of ethanol (96-100%) was added.

496   The samples were vortexed at maximum speed for 30 sec, added to a DNeasy Mini spin column,

497   centrifuged (16,000 x g, 1 min), and the supernatant was discarded. Buffer AW2 (Qigen) was

498   added (500 µl), followed by centrifugation (20,000 x g, 3 min). The DNeasy Mini spin column was

499   placed into a sterile 2 ml microcentrifuge tube, and the gDNA was eluted in 100 µl of AE Buffer

500   by centrifugation (20,000 x g, 1 min) following a 1 min incubation at room temperature. The

501   gDNA concentration was quantified (1 µl sample volume) with a Thermo Scientific™ NanoDrop™

502   One Microvolume UV-Vis Spectrophotometer (840274100) and stored at -20°C.

503

504   *Bacterial Genomic DNA extraction for PacBio whole-genome sequencing*

505   High molecular weight genomic DNA (20-160 kb) was extracted from four Streptomyces strains

506   (DF, SFW, QF2, and JS) with the QIAGEN-Genomic-tip 500/G kit (10262) according to the

507   manufacturer's protocol for bacteria.

508

509   *Bacterial whole-genome sequencing, assembly, and annotation*

510   The genome sequences of four Streptomyces strains were generated using the Pacific

511   Biosciences RS II (PacBio RS II) single molecule real-time (SMRT) sequencing platform at the

512   IGM Genomics Center, University of California, San Diego, La Jolla, CA. Genome sequences

513   were assembled using the HGAP protocol integrated in the PacBio RS II sequencer (smrt

514   analysis v2.3.0/Patch5) resulting in a variable number (n = 1-95) of contigs per genome, and

515   ranged in size from 5.42 to 7.79 Mb. The mauve contig mover was used to order the contigs of

516   three draft genome sequences (genomes of strains SFW, QF2, and JS) relative to a closely

517   related reference sequence (*S. pratensis* ATCC 33331, *S. globisporus* C-1027, and *S. parvulus*

518   2297 respectively). DNA sequencing of strain DF resulted in a single contig and did not require

519   reordering to restore gene synteny. Gene prediction and annotation were made with the Rapid

520   Annotations using Subsystems Technology (RASTtk) platform.[70]

521

522   *Phage genomic DNA extraction*

523   5 µl of RNase A and 5 µl of DNase I were added to 10ml of lysate, incubated at 30°C for 30

524   minutes, and then precipitated overnight at 4°C by the addition of 4 ml of 20% polyethylene

525   glycol 8,000. Samples were centrifuged at 10,000 g's for 30 minutes, and pellets resuspended

526   in Qiagen PB buffer and DNA isolated using a Qiagen plasmid DNA isolation column as

527   recommended by the manufacturer.

528

529   *Phage genome sequencing, assembly, and annotation*

530   Genomic DNA of 4 bacteriophages (TrvxScott, BartholomewSD, Shawty, and IceWarrior) was

531   sequenced using the Illumina MiSeq platform at the Pittsburgh Bacteriophage Institute

532   sequencing facility. The genomes were assembled with Newbler and checked for quality with

533   Consed. The whole genome sequences were submitted to GenBank (Acc No. MH669016,

534   MK460245, MK433266, and MK433259). DNA Master was used for annotation, and NCBI

535   BLASTp was used to determine the potential function of gene products. Whole genome

536   sequence comparisons were performed in Phamerator.[71]

537

538   *16S rRNA PCR amplification and sequencing*

539   16S ribosomal DNA templates (~1,465 bp) were amplified using Q5 high fidelity PCR (New

540   England Biolabs) with the universal primer set 27F (5'-AGAGTTTGATCCTGGCTCAG-3') and

541   1492R (5'-GGTTACCTTGTTACGACTT-3').[72] Each PCR mixture (50 µl) contained 100 ng of

542   template gDNA, 500 pmol of each primer, and 200 µM dNTPs. PCR thermocycling conditions

543   were as follows: 30 seconds of initial denaturation at 98°C, 30 cycles of denaturation at 98°C for

544   10 seconds, annealing for 15 seconds at 60°C, extension at 72°C for 1.5 minutes, and a final

545   extension at 72°C for 5 minutes then held at 4°C. PCR products were purified with the

546   oligonucleotide cleanup protocol as described in the Monarch PCR & DNA Cleanup Kit 5 µg user

547   manual (NEB #T1030). Clean PCR products were sequenced using Sanger methods by Eton

548   Biosciences (https://www.etonbio.com/) and trimmed for quality before analysis.

549

550   *CRISPR-Cas sequence analysis and predictions*

551   The sequences of all four Streptomyces were searched for CRISPR arrays (repeats and spacers)

552   and potentially associated Cas genes using the following software tools; CRISPR-Cas++,[64, 65]

553   CRISPROne,[66] CRISPRMiner2,[67] and CRISPRDetect.[68]

554

555   *Phylogenetic analyses of bacterial isolates*

556   16S rRNA sequences were trimmed on both ends, (5' and 3') in Geneious Prime using the Trim

557   Ends function with an error probability limit set at 0.05, which trims regions with more than a 5%

558   chance of an error per base. Sequences were aligned using MUSCLE v3.8.425 with a maximum

559   of 1,000 iterations, and profile-dependent parameters were set. Sequences were grouped

560   based on similarity, and anchor optimized. Distance was measured using kmer4_6 for the first

561   iteration, and subsequent distance measurements were performed with pctid_kimura. For all

562   iterations the clustering method was neighbor joining, and the sequence weighting scheme was

563   ClustalW with an spm objective score, 32 anchor spacing and -1 gap open score. Diagonals

564   were set at a minimum length of 24 and a margin of 5, the minimum column anchr scores were

565   set min best and min smoothed at 90. Finally, the hydrophobicity multiplier was set at 1.2 with

566   a window size of 5. The tree was created with the Geneious Tree Builder using the Jukes-Cantor

567   genetic distance model, Neighbor-Joining method to build the tree, with the outgroup set as

568   *Chlorobium limicola* strain DSM 245. The resulting consensus tree was obtained after 1,000

569   iterations of bootstrap resampling, with a support threshold set at 70% and a random seed of

570   409,923.

571

572   *Cross-streak method for assessing antibacterial production potential*

573   From a single colony, using sterile Q-tips, Streptomyces isolates were streaked in a broad

574   vertical line (2 inch) onto LB, and AIA, solid agar plates and incubated for one week at 30°C. The

575   day before the assay, test strains (*E. coli* JP313 Δ*tolC*, *B. subtilis* PY79, and *E. coli* MC4100)

576   were grown in 5 ml of LB and incubated at 30°C overnight while rolling. On the day of the

577   antibacterial screen, the overnight cultures of each test strain were diluted (1:100 in 5 ml LB) and

578   grown to log phase OD_{600} 0.15-0.2 (~1.5 hr at 30°C while rolling). A volume of 10 μl of each test

579   strain was spotted in distinct lines almost to the edge of the Streptomyces line at a perpendicular

580   angle. The plates were incubated overnight at 30°C, then investigated for the presence of zones

581   of inhibition which were measured in millimeters.

582

583   *Bacterial cytological profiling (BCP) on plates*

584   Fluorescence microscopy and BCP on plates was performed as previously described by

585   Nonejuie et. al.[15] Briefly, Streptomyces strains (DF, SFW, QF2, and JS) were streaked in a vertical

586   line down the center of LB, AIA, and ISP2 plates (for one liter: 4.0 g Difco yeast extract, 10.0 g

587   Difco malt extract, 4.0 g dextrose, 20.0 g agar, pH 7.0), incubated for one week at 30°C. The

588   test strain, *E. coli* JP313 Δ*tolC*, was prepared and spotted as described above in the cross-

589   streak method. Following a 2 hr incubation at 30°C, a 1.5 x 1.5 cm square (~2.5 cm^2) piece of

590   agar containing the *E. coli* test strain was cut and prepared for high resolution fluorescence

591   microscopy. The cut piece of agar was placed on a microscope slide, the *E. coli* cells were

592   stained with fluorescent dyes, a coverslip was placed on top of the stained cells then imaged.

593

594   *Host range experiment*

595   The host ranges of 4 phages were determined against the Streptomyces strains: QF2, DF, JS,

596   and SFW*.* The experiment was blinded by assigning phages numbers i-iv and hosts letters A-D.

597   A lawn of Streptomyces in LB top agar was poured on LB CHX plates. After the top agar

598   solidified, a grid was drawn on the bottom of the plate, and 5 µl of pre-diluted phage samples

599   (10^0 to 10^{-10} in phage buffer) were spotted in squares on the grid. Plaques were counted and

600   used to calculate a titer, which was then compared to the titer obtained against *S. platensis* in

601   order to calculate the efficiency of infection.

602

603   *Transmission electron microscopy*

604   10 µl of lysate was applied to a copper grid, stained with 1% uranyl acetate, washed twice with

605   phage buffer, and allowed to dry. Images were collected using a FEI Tecnai Spirit G2 BioTWIN

606   Transmission Electron Microscope equipped with a bottom mount Eagle 4k camera.

607

608   *Strains used in this study*

609   *We used the following strains: S. platensis JCM 4664 substr. MJ1A1, E. coli MC4100, B. subtilis*

610   PY79, *P. aeruginosa* P4, *S. aureus* MRSA USA300 TCH1516 from Texas Children's Hospital

611   (USA300-HOU-MR), *S. coelicolor* A3(2) substr. M146, *E. coli* JP313 *ΔtolC*, as well as two strains

612   generously donated by Prof. Keith Poole at Queens University in Kingston, Canada – *P.*

613   *aeruginosa* PA01 and *P. aeruginosa* K2733 Δefflux (*ΔMexAB–OprM, ΔMexCD–OprJ, ΔMexEF–*

614   *OprN, ΔMexXY–OprM*). The Δ*tolC5* mutation is derived from strain EW1b (CGSC #5634), and

615   was introduced into strain JP313 by P1 transduction. JP313 was transduced to tetracycline

616   resistance with a lysate of strain CAG18475 (*metC162*::Tn*10*), and the methionine requirement

617   of the transductants was confirmed. This strain was then transduced to prototrophy with a lysate

618   of EW1b, and these transductants were screened on MacConkey agar for the presence of the

619   *ΔtolC5* mutation. EW1b and CAG18475 were obtained from the *Coli* Genetic Stock Center at

620   Yale University.

Data Availability

621

622   All data generated or analyzed in this study are included in this article and its supplementary

623   information files. Nucleotide sequences of the phages in this study have been deposited in the

624   GenBank database with accession numbers: MK460245 (BartholomewSD), MH669016

625   (TrvxScott), MK433266 (Shawty), and MK433259 (IceWarrior).

626

Competing Interests

627

628   The authors declare no competing interests.

629

References

630

631   1.   Wencewicz, T.A., *Crossroads of Antibiotic Resistance and Biosynthesis.* J Mol Biol, 2019.
632   **431**(18): p. 3370-3399.
633   2.   Pogue, J., et al., *Appropriate antimicrobial therapy in the era of multidrug-resistant human*
634   *pathogens.* Clinical Microbiology and Infection, 2015. **21**(4): p. 302-312.
635   3.   Brown, E.D. and G.D. Wright, *Antibacterial drug discovery in the resistance era.* Nature,
636   2016. **529**(7586): p. 336-343.
637   4.   Newman, D.J. and G.M. Cragg, *Natural products as sources of new drugs from 1981 to*
638   *2014.* Journal of natural products, 2016. **79**(3): p. 629-661.
639   5.   Fischbach, M.A. and C.T. Walsh, *Antibiotics for emerging pathogens.* Science, 2009.
640   **325**(5944): p. 1089-1093.
641   6.   Hover, B.M., et al., *Culture-independent discovery of the malacidins as calcium-*
642   *dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens.*
643   Nature microbiology, 2018. **3**(4): p. 415-422.
644   7.   Ling, L.L., et al., *A new antibiotic kills pathogens without detectable resistance.* Nature,
645   2015. **517**(7535): p. 455-459.
646   8.   Chen, R., H.L. Wong, and B.P. Burns, *New approaches to detect biosynthetic gene*
647   *clusters in the environment.* Medicines, 2019. **6**(1): p. 32.
648   9.   Pye, C.R., et al., *Retrospective analysis of natural products provides insights for future*
649   *discovery trends.* Proceedings of the National Academy of Sciences, 2017. **114**(22): p.
650   5601-5606.
651   10.   Cox, G., et al., *A Common Platform for Antibiotic Dereplication and Adjuvant Discovery.*
652   Cell Chem Biol, 2017. **24**(1): p. 98-109.
653   11.   Nonejuie, P., et al., *Bacterial cytological profiling rapidly identifies the cellular pathways*
654   *targeted by antibacterial molecules.* Proceedings of the National Academy of Sciences,
655   2013. **110**(40): p. 16169-16174.

656   12.   Lamsa, A., et al., *Rapid inhibition profiling in Bacillus subtilis to identify the mechanism
657   of action of new antimicrobials.* ACS chemical biology, 2016. **11**(8): p. 2222-2231.
658   13.   Lamsa, A., et al., *The Bacillus subtilis cannibalism toxin SDP collapses the proton motive
659   force and induces autolysis.* Molecular microbiology, 2012. **84**(3): p. 486-500.
660   14.   Peters, C.E., et al., *Rapid Inhibition Profiling Identifies a Keystone Target in the Nucleotide
661   Biosynthesis Pathway.* ACS chemical biology, 2018. **13**(12): p. 3251-3258.
662   15.   Nonejuie, P., et al., *Application of bacterial cytological profiling to crude natural product
663   extracts reveals the antibacterial arsenal of Bacillus subtilis.* The Journal of antibiotics,
664   2016. **69**(5): p. 353.
665   16.   Quach, D., et al., *Bacterial cytological profiling (BCP) as a rapid and accurate
666   antimicrobial susceptibility testing method for Staphylococcus aureus.* EBioMedicine,
667   2016. **4**: p. 95-103.
668   17.   Alberti, F. and C. Corre, *Editing streptomycete genomes in the CRISPR/Cas9 age.* Natural
669   product reports, 2019. **36**(9): p. 1237-1248.
670   18.   de Jonge, P.A., et al., *Molecular and evolutionary determinants of bacteriophage host
671   range.* Trends in microbiology, 2019. **27**(1): p. 51-63.
672   19.   Russell, D.A. and G.F. Hatfull, *PhagesDB: the actinobacteriophage database.*
673   Bioinformatics, 2017. **33**(5): p. 784-786.
674   20.   Genilloud, O., *Actinomycetes: still a source of novel antibiotics.* Natural product reports,
675   2017. **34**(10): p. 1203-1232.
676   21.   Thorpe, H.M., S.E. Wilson, and M.C. Smith, *Control of directionality in the site-specific
677   recombination system of the Streptomyces phage φC31.* Molecular microbiology, 2000.
678   **38**(2): p. 232-241.
679   22.   Fogg, P.C.M., et al., *New Applications for Phage Integrases.* Journal of Molecular
680   Biology, 2014. **426**(15): p. 2703-2716.
681   23.   Morita, K., et al., *The site-specific recombination system of actinophage TG1.* FEMS
682   Microbiol Lett, 2009. **297**(2): p. 234-40.
683   24.   Morita, K., et al., *Site-specific genome integration in alphaproteobacteria mediated by
684   TG1 integrase.* Applied Microbiology and Biotechnology, 2012. **93**(1): p. 295-304.
685   25.   Venken, K.J.T., et al., *Genome engineering: Drosophila melanogaster and beyond.* Wiley
686   interdisciplinary reviews. Developmental biology, 2016. **5**(2): p. 233-267.
687   26.   Thakur, D., et al., *Influence of nutrition and culturing conditions for optimum growth and
688   antimicrobial metabolite production by Streptomyces sp. 201.* Journal de Mycologie
689   Medicale, 2009. **19**(3): p. 161-167.
690   27.   Zimmerman, S.B., *Toroidal nucleoids in Escherichia coli exposed to chloramphenicol.*
691   Journal of structural biology, 2002. **138**(3): p. 199-206.
692   28.   Aziz, R.K., et al., *The RAST Server: rapid annotations using subsystems technology.* BMC
693   genomics, 2008. **9**(1): p. 75.
694   29.   Blin, K., et al., *antiSMASH 5.0: updates to the secondary metabolite genome mining
695   pipeline.* Nucleic acids research, 2019. **47**(W1): p. W81-W87.
696   30.   Kautsar, S.A., et al., *MIBiG 2.0: a repository for biosynthetic gene clusters of known
697   function.* Nucleic acids research, 2020. **48**(D1): p. D454-D458.
698   31.   Holm, M., et al., *Molecular mechanism of viomycin inhibition of peptide elongation in
699   bacteria.* Proceedings of the National Academy of Sciences, 2016. **113**(4): p. 978-983.
700   32.   Komatsu, Y. and K. Tanaka, *Mechanism of Action of Showdomycin: Part II. Effect of
701   Showdomycin on the Synthesis of Deoxyribonucleic Acid in Escherichia coli.* Agricultural
702   and Biological Chemistry, 1970. **34**(6): p. 891-899.
703   33.   Reusser, F., *Ficellomycin and feldamycin; inhibitors of bacterial semiconservative DNA
704   replication.* Biochemistry, 1977. **16**(15): p. 3406-3412.

705 34. Pohle, S., et al., *Biosynthetic gene cluster of the non-ribosomally synthesized*
706   *cyclodepsipeptide skyllamycin: deciphering unprecedented ways of unusual*
707   *hydroxylation reactions.* Journal of the American Chemical Society, 2011. **133**(16): p.
708   6194-6205.
709 35. Kanda, N., et al., *DB-2073, a new alkylresorcinol antibiotic.* The Journal of antibiotics,
710   1975. **28**(12): p. 935-942.
711 36. GÜRTLER, H., et al., *Albaflavenone, a sesquiterpene ketone with a zizaene skeleton*
712   *produced by a streptomycete with a new rope morphology.* The Journal of antibiotics,
713   1994. **47**(4): p. 434-439.
714 37. Zhao, B., et al., *Biosynthesis of the sesquiterpene antibiotic albaflavenone in*
715   *Streptomyces coelicolor A3 (2).* Journal of Biological Chemistry, 2008. **283**(13): p. 8183-
716   8189.
717 38. Hobbs, G., et al., *An integrated approach to studying regulation of production of the*
718   *antibiotic methylenomycin by Streptomyces coelicolor A3 (2).* Journal of bacteriology,
719   1992. **174**(5): p. 1487-1494.
720 39. Baltz, R.H., V. Miao, and S.K. Wrigley, *Natural products to drugs: daptomycin and related*
721   *lipopeptide antibiotics.* Natural Product Reports, 2005. **22**(6): p. 717-741.
722 40. Drautz, H., W. Keller-Schierlein, and H. Zähner, *Stoffwechselprodukte von*
723   *Mikroorganismen.* Archives of Microbiology, 1975. **106**(3): p. 175-190.
724 41. Ballard, T.E. and C. Melander, *Kinamycin-mediated DNA cleavage under biomimetic*
725   *conditions.* Tetrahedron Letters, 2008. **49**(19): p. 3157-3161.
726 42. Nagarajan, R., et al., *β-Lactam Antibiotics from Streptomyces.* Journal of the American
727   Chemical Society, 1971. **93**(9): p. 2308-2310.
728 43. Papp-Wallace, K.M., et al., *Carbapenems: past, present, and future.* Antimicrobial agents
729   and chemotherapy, 2011. **55**(11): p. 4943-4960.
730 44. Neu, H.C. and K.P. Fu, *Clavulanic acid, a novel inhibitor of β-lactamases.* Antimicrobial
731   agents and chemotherapy, 1978. **14**(5): p. 650-655.
732 45. Suzuki, J., T. KUNIMOTO, and M. HORI, *Effects of kanamycin on protein synthesis:*
733   *inhibition of elongation of peptide chains.* The Journal of antibiotics, 1970. **23**(2): p. 99-
734   101.
735 46. Zhang, X., et al., *Biosynthetic Investigations of Lactonamycin and Lactonamycin Z:*
736   *Cloning of the Biosynthetic Gene Clusters and Discovery of an Unusual Starter Unit.*
737   Antimicrobial Agents and Chemotherapy, 2008. **52**(2): p. 574.
738 47. Matsumoto, N., et al., *Lactonamycin, a new antimicrobial antibiotic produced by*
739   *Streptomyces rishiriensis MJ773-88K4.* The Journal of antibiotics, 1999. **52**(3): p. 269-
740   275.
741 48. Yamamoto, H., et al., *Ribosomal resistance of an istamycin producer,*
742   *Streptomycestenjimariensis, to aminoglycoside antibiotics.* Biochemical and biophysical
743   research communications, 1981. **100**(3): p. 1396-1401.
744 49. Hoang, N.H., et al., *Istamycin aminoglycosides profiling and their characterization in*
745   *Streptomyces tenjimariensis ATCC 31603 culture using high-performance liquid*
746   *chromatography with tandem mass spectrometry.* Journal of separation science, 2016.
747   **39**(24): p. 4712-4722.
748 50. Okami, Y., et al., *New aminoglycoside antibiotics, istamycins A and B.* The Journal of
749   antibiotics, 1979. **32**(9): p. 964-966.
750 51. Hertweck, C., et al., *Type II polyketide synthases: gaining a deeper insight into enzymatic*
751   *teamwork.* Natural product reports, 2007. **24**(1): p. 162-190.
752 52. Newbold, C.J., et al., *Effect of the novel ionophore tetronasin (ICI 139603) on ruminal*
753   *microorganisms.* Appl. Environ. Microbiol., 1988. **54**(2): p. 544-547.

754   53.   Kusche, B.R., et al., *Alternating pattern of stereochemistry in the nonactin macrocycle is required for antibacterial activity and efficient ion binding.* Journal of the American Chemical Society, 2009. **131**(47): p. 17155-17165.

757   54.   Cruz-Morales, P., et al., *Phylogenomic analysis of natural products biosynthetic gene clusters allows discovery of arseno-organic metabolites in model streptomycetes.* Genome biology and evolution, 2016. **8**(6): p. 1906-1916.

760   55.   Wang, L.-H., et al., *Membrane and genomic DNA dual-targeting of citrus flavonoid naringenin against Staphylococcus aureus.* Integrative Biology, 2017. **9**(10): p. 820-829.

762   56.   Kleijn, L.H., et al., *Total synthesis of Laspartomycin C and characterization of its antibacterial mechanism of action.* Journal of medicinal chemistry, 2016. **59**(7): p. 3569-3574.

765   57.   Rui, Z., et al., *Biochemical and genetic insights into asukamycin biosynthesis.* Journal of Biological Chemistry, 2010. **285**(32): p. 24915-24924.

767   58.   Li, C., X. Li, and R. Hong, *Synthetic Study on Tetrapetalones: Stereoselective Cyclization of N-Acyliminium Ion To Construct Substituted 1-Benzazepines.* Organic letters, 2009. **11**(18): p. 4036-4039.

770   59.   Luo, Y., et al., *Activation and characterization of a cryptic polycyclic tetramate macrolactam biosynthetic gene cluster.* Nature communications, 2013. **4**(1): p. 1-8.

772   60.   Smanski, M.J., et al., *Engineered Streptomyces platensis Strains That Overproduce Antibiotics Platensimycin and Platencin.* Antimicrobial Agents and Chemotherapy, 2009. **53**(4): p. 1299.

775   61.   Burke, J., D. Schneider, and J. Westpheling, *Generalized transduction in Streptomyces coelicolor.* Proceedings of the National Academy of Sciences, 2001. **98**(11): p. 6289.

778   62.   Ackermann, H.-W., *5500 Phages examined in the electron microscope.* Archives of virology, 2007. **152**(2): p. 227-243.

780   63.   Jin, H., et al., *Capsid Structure of a Freshwater Cyanophage Siphoviridae Mic1.* Structure, 2019. **27**(10): p. 1508-1516. e3.

782   64.   Couvin, D., et al., *CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins.* Nucleic acids research, 2018. **46**(W1): p. W246-W251.

785   65.   Abby, S.S., et al., *MacSyFinder: a program to mine genomes for molecular systems with an application to CRISPR-Cas systems.* PloS one, 2014. **9**(10).

787   66.   Zhang, Q. and Y. Ye, *Not all predicted CRISPR–Cas systems are equal: isolated cas genes and classes of CRISPR like elements.* BMC bioinformatics, 2017. **18**(1): p. 92.

789   67.   Zhang, F., et al., *CRISPRminer is a knowledge base for exploring CRISPR-Cas systems in microbe and phage interactions.* Communications Biology, 2018. **1**(1): p. 1-5.

791   68.   Biswas, A., et al., *CRISPRDetect: a flexible algorithm to define CRISPR arrays.* BMC genomics, 2016. **17**(1): p. 356.

793   69.   Faure, G., et al., *CRISPR–Cas in mobile genetic elements: counter-defence and beyond.* Nature Reviews Microbiology, 2019. **17**(8): p. 513-525.

795   70.   Brettin, T., et al., *RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes.* Scientific Reports, 2015. **5**(1): p. 8365.

798   71.   Cresawn, S.G., et al., *Phamerator: a bioinformatic tool for comparative bacteriophage genomics.* BMC Bioinformatics, 2011. **12**(1): p. 395.

800   72.   Janda, J.M. and S.L. Abbott, *16S rRNA Gene Sequencing for Bacterial Identification in the Diagnostic Laboratory: Pluses, Perils, and Pitfalls.* Journal of Clinical Microbiology, 2007. **45**(9): p. 2761.
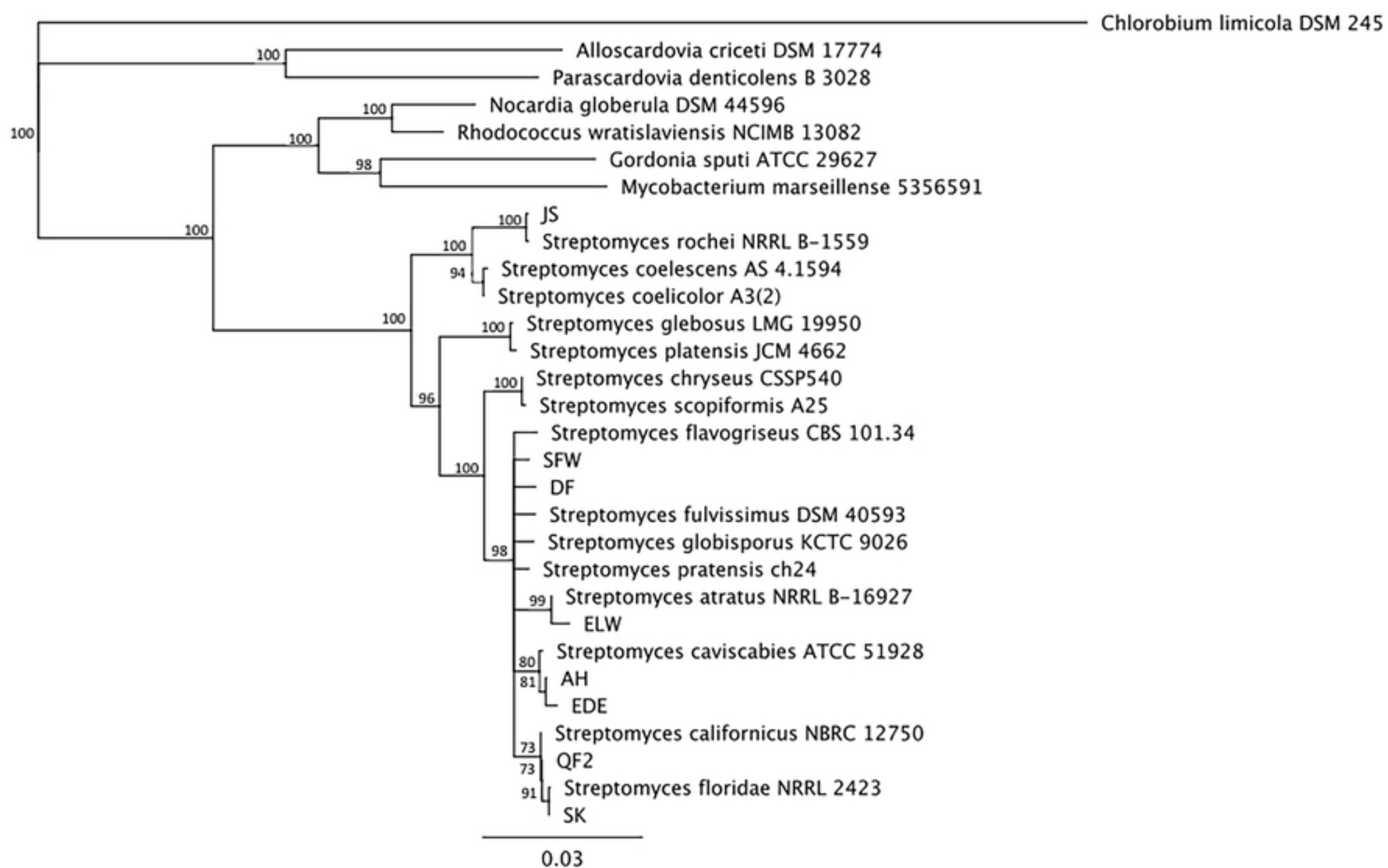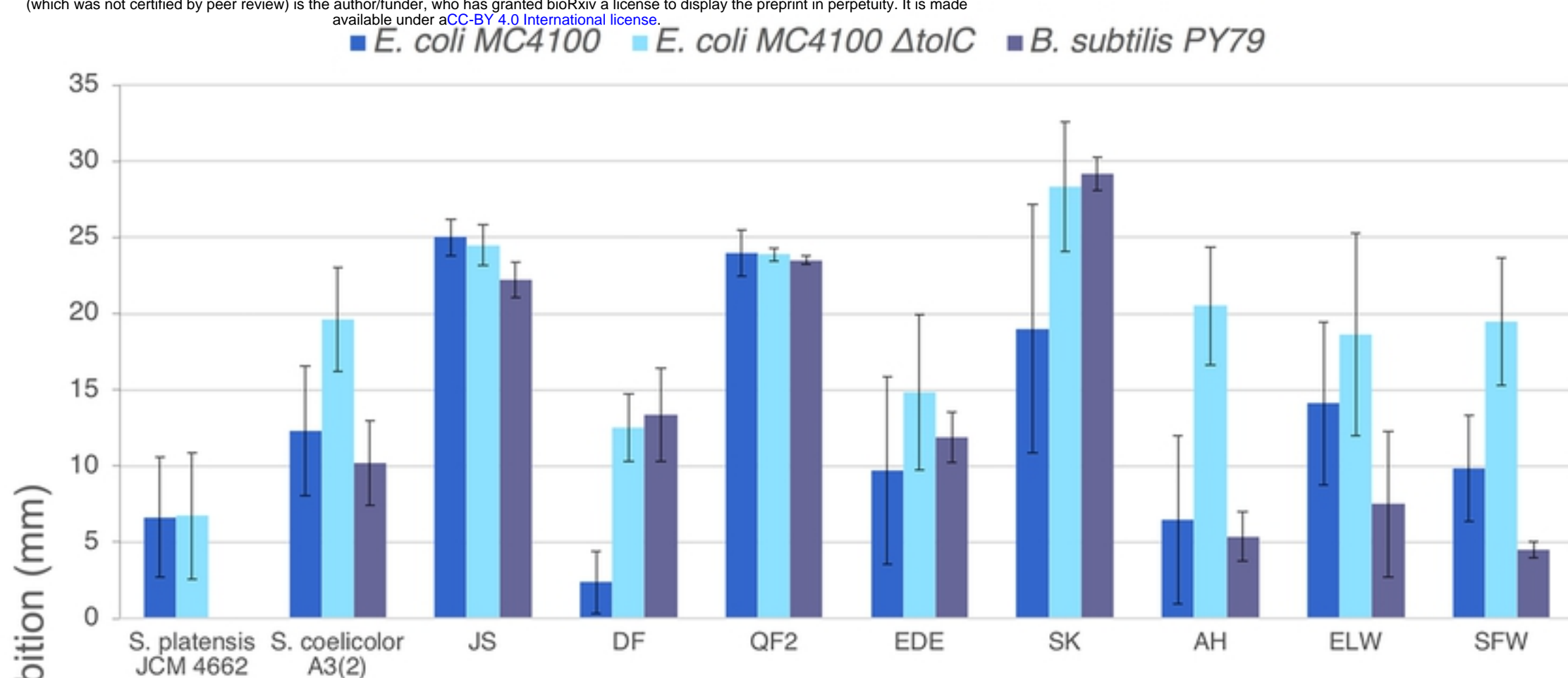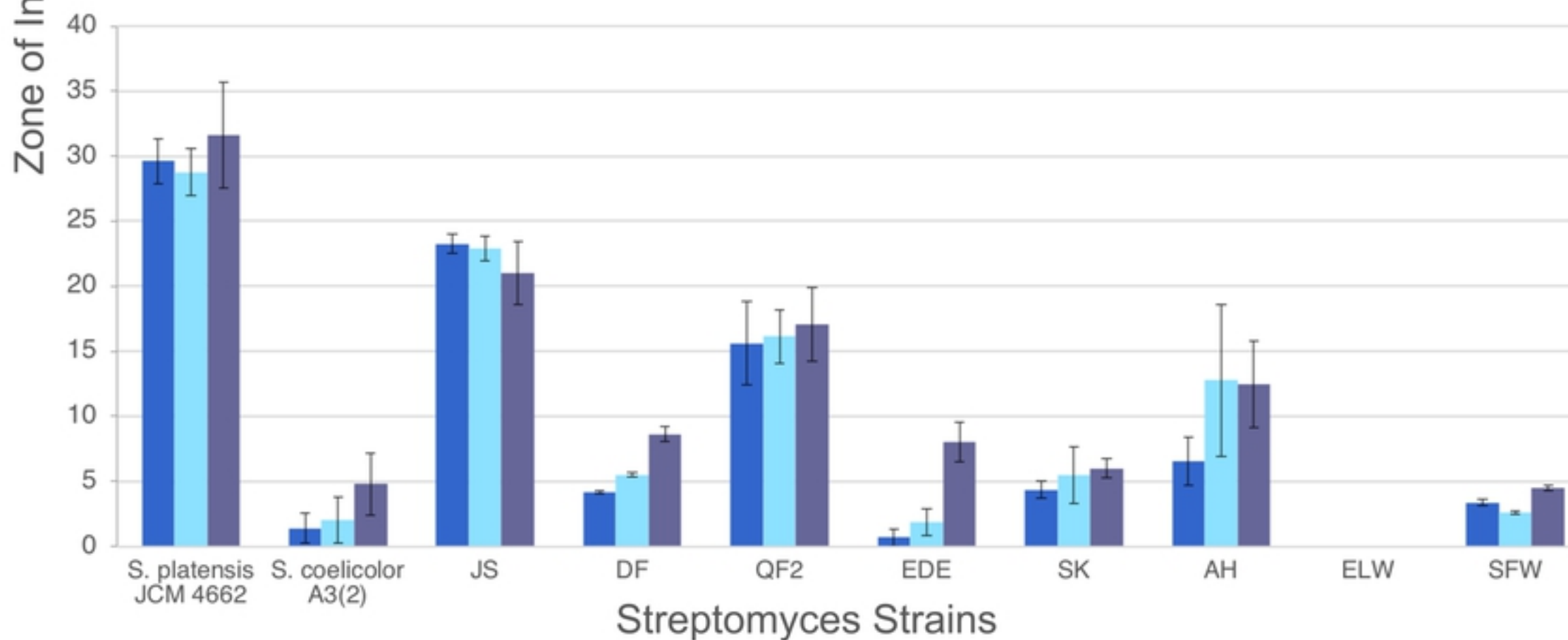
803

Chlorobium limicola DSM 245

Alloscardovia criceti DSM 17774
Parascardovia denticolens B 3028
Nocardia globerula DSM 44596
Rhodococcus wratislaviensis NCIMB 13082
Gordonia sputi ATCC 29627
Mycobacterium marseillense 5356591
JS
Streptomyces rochei NRRL B-1559
Streptomyces coelescens AS 4.1594
Streptomyces coelicolor A3(2)
Streptomyces glebosus LMG 19950
Streptomyces platensis JCM 4662
Streptomyces chryseus CSSP540
Streptomyces scopiformis A25
Streptomyces flavogriseus CBS 101.34
SFW
DF
Streptomyces fulvissimus DSM 40593
Streptomyces globisporus KCTC 9026
Streptomyces pratensis ch24
Streptomyces atratus NRRL B-16927
ELW
Streptomyces caviscabies ATCC 51928
AH
EDE
Streptomyces californicus NBRC 12750
QF2
Streptomyces floridae NRRL 2423
SK

0.03

Figure 1

Figure 2

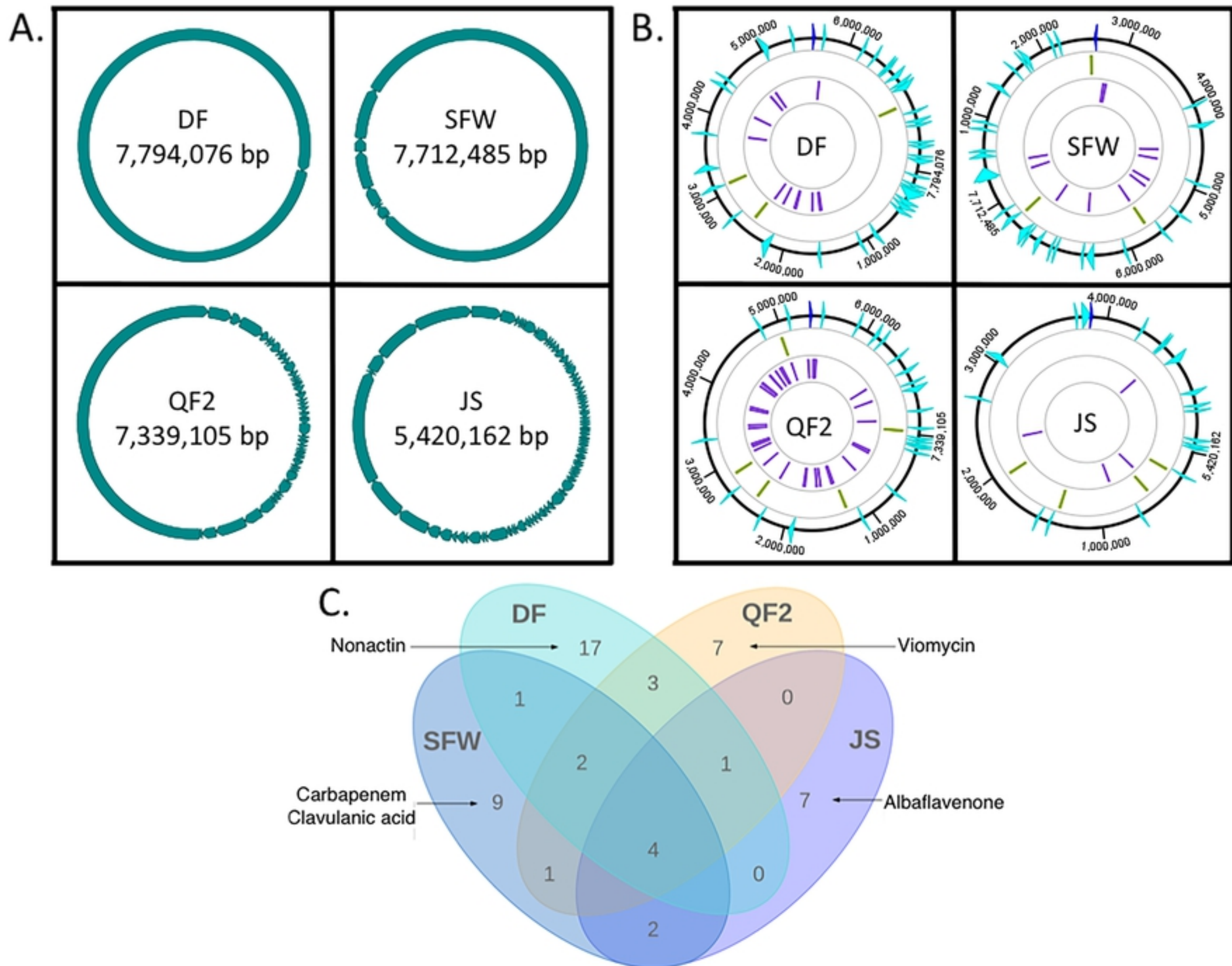|	LB	AIA	ISP2
E. coli MC4100 Δ*tolC*			
DF			
SFW			
QF2			
JS			

Figure 3

Figure 4

Figure 5

A.

TrvxScott   BartholomewSD   Shawty   IceWarrior

100 nm

B.

TrvxScott (BD2)

BartholomewSD (BD2)

Shawty (BB1)

IceWarrior (BI1)
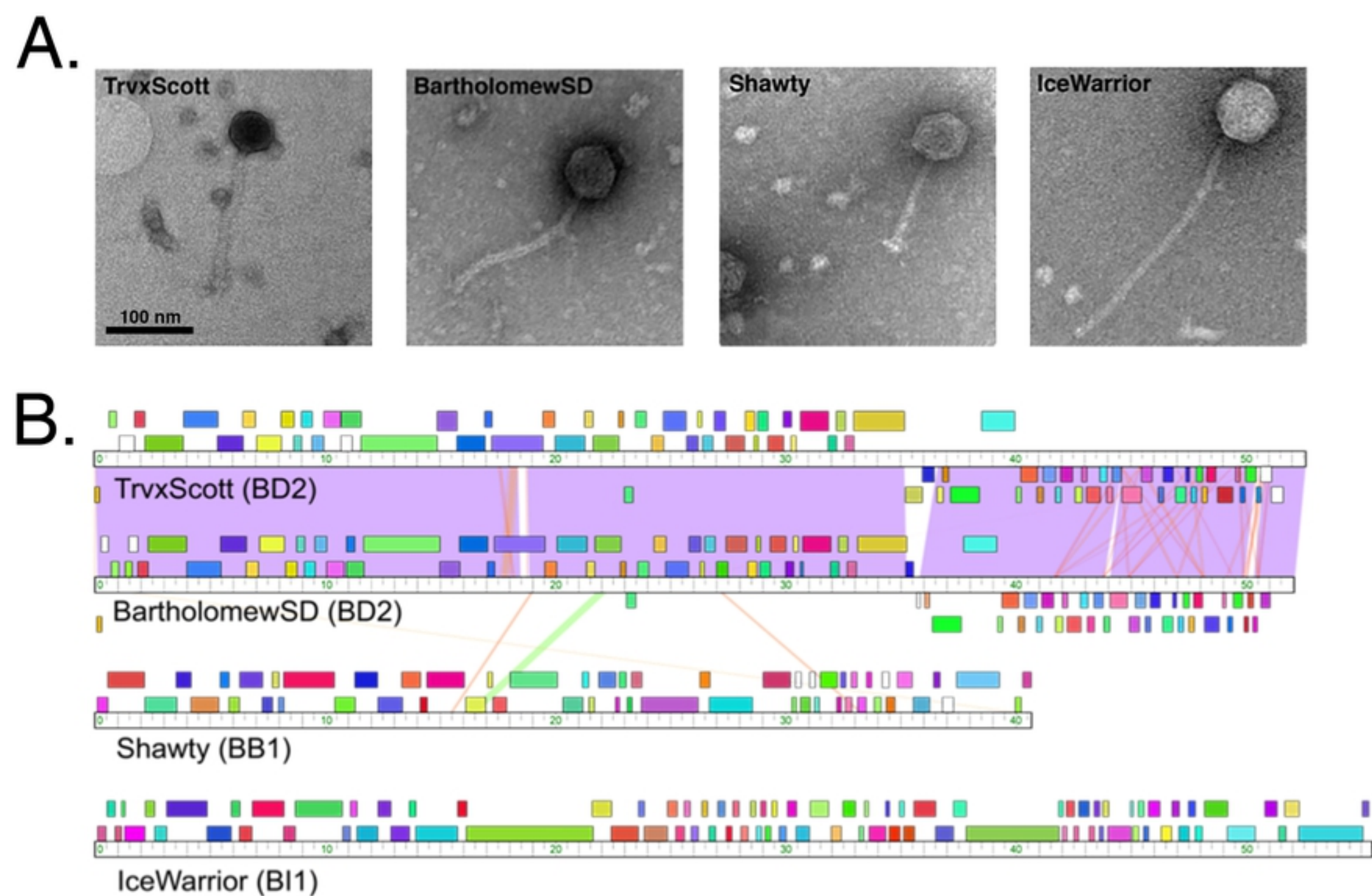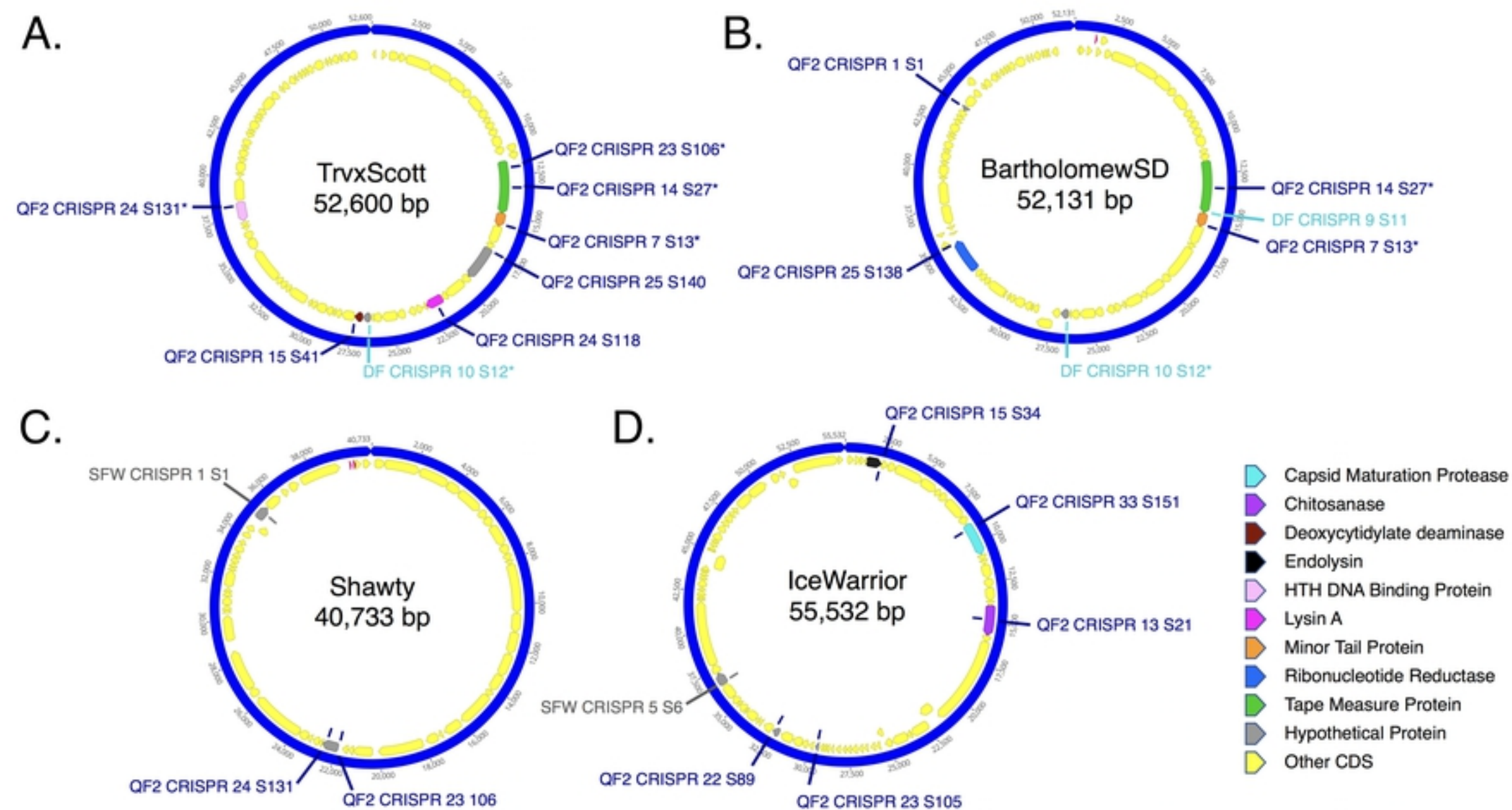
Figure 6

Figure 7

Type I-E CRISPR-Cas System

Strain QF2

4,973,482

6.4 Kb   4,987,810

Cas3 HD   Cas3"   Cas8e   Cse2   Cas7   Cas5e Cas6   CRISPR 31

E. coli K-12 MG1655

2,887,219

2,877,618

Cas3"   Cas8e   Cse2   Cas7   Cas5e Cas6   Cas1 Cas2   CRISPR

Figure 8