

# Supplementary text 3:

## Protospacer examples for selected genera

<u>organism</u>		<u>Escherichia coli strain C15</u>							
Arrays found		Start	Stop	Size	Repeat	Spacers	CRISPR Type	Organism	
CP011018.1		4417105		4418353	1248	GTGTTCCCC	20 Unidentified	Escherichia_coli_C15_GCA_000971615.1	
CP011018.1		4444052		4445606	1554	GTGTTCCCC	25 CAS-I-E	Escherichia_coli_C15_GCA_000971615.1	
Hits		ProtospacerTy	ProtospacerContig	Protospace	ProtospaceCRISPR Typ	CRISPR	Organism	ProtospaceTarget	ann Note
SpacerID									
CP011018.1_4444052_4445606_16_spacer_4444996_33	Phage	LRXJ01000139		68082	68114	CAS-I-E	Escherichia_coli_C15_GCA_000971615.1	562	562 Escherichia coli strain LV2795 Unknown Prophage region
CP011018.1_4444052_4445606_17_spacer_4445058_32	Phage	LRXJ01000139		65460	65491	CAS-I-E	Escherichia_coli_C15_GCA_000971615.1	562	562 Escherichia coli strain LV2795 Unknown Prophage region
CP011018.1_4444052_4445606_19_spacer_4445180_32	Phage	AFDY01000015		187752	187783	CAS-I-E	Escherichia_coli_C15_GCA_000971615.1	562	562 Escherichia coli STEC_H.1.8 geUnknown Prophage region
CP011018.1_4444052_4445606_22_spacer_4445363_32	Phage	AFDY01000015		168382	168413	CAS-I-E	Escherichia_coli_C15_GCA_000971615.1	562	562 Escherichia coli STEC_H.1.8 gepHage N-6-adenine-met Prophage region
CP011018.1_4444052_4445606_3_spacer_4444203_32	Phage	BBVG01000009		206444	206475	CAS-I-E	Escherichia_coli_C15_GCA_000971615.1	562	562 Escherichia albertii DNA, cont Unknown
CP011018.1_4444052_4445606_6_spacer_4444386_32	Phage	AQG001000031		45042	45073	CAS-I-E	Escherichia_coli_C15_GCA_000971615.1	562	562 Escherichia coli STEC O174:H2Unknown
CP011018.1_4444052_4445606_9_spacer_4444569_32	Phage	KB732762		2420768	2420799	CAS-I-E	Escherichia_coli_C15_GCA_000971615.1	562	562 Escherichia coli KTE66 genom Unknown Prophage region

<u>Organism</u>		<u>Sulfolobus solfataricus P2</u>									
Arrays found		Start	Stop	Size	Repeat	Spacers	CRISPR Type	Organism			
AE006641.1		1809772		1815557	5785	CTTTCAATT	88 CAS-III-B	Sulfolobus_solfataricus_P2_GCA_000007005.1			
AE006641.1		1809772		1815557	5785	CTTTCAATT	88 CAS-I-A	Sulfolobus_solfataricus_P2_GCA_000007005.1			
AE006641.1		1233466		1239959	6493	GATTAATCC	102 Unidentified	Sulfolobus_solfataricus_P2_GCA_000007005.1			
AE006641.1		1254482		1260452	5970	CTTTCAATT	94 CAS-I-A	Sulfolobus_solfataricus_P2_GCA_000007005.1			
AE006641.1		1297153		1299148	1995	CTTTCAATT	31 CAS-I-A	Sulfolobus_solfataricus_P2_GCA_000007005.1			
AE006641.1		1305634		1311637	6003	GATAATCTC	95 CAS-I-A	Sulfolobus_solfataricus_P2_GCA_000007005.1			
AE006641.1		1744007		1744417	410	GATAATCTA	6 Unidentified	Sulfolobus_solfataricus_P2_GCA_000007005.1			
SpacerID		Protospace	ProtospacerCont	Protospace	ProtospaceCRISPR Typ	CRISPR	Organism	Protospacer	Organism	Target	ann Note
AE006641.1_1305634_1311637_24_spacer_1307110_38	Phage	HG322870.1		1473	1510	CAS-I-A	Sulfolobus_solfataricus_P2_GCA_00Sulfolobus monocaudavirus SMV1 complete genome			conserved	archaeal viral membrane protein
AE006641.1_1305634_1311637_26_spacer_1307235_38	Phage	HG322870.1		46684	46721	CAS-I-A	Sulfolobus_solfataricus_P2_GCA_00Sulfolobus monocaudavirus SMV1 complete genome			Conserved	archaeal viral membrane protein, acyltransferase family
AE006641.1_1305634_1311637_28_spacer_1307363_39	Phage	KX607102.1		19586	19548	CAS-I-A	Sulfolobus_solfataricus_P2_GCA_00Acidianus two-tailed virus variant 1 isolate ATV.v1, complete genome			archaeal	structural protein
AE006641.1_1305634_1311637_29_spacer_1307426_40	Phage	KX607102.1		50321	50360	CAS-I-A	Sulfolobus_solfataricus_P2_GCA_00Acidianus two-tailed virus variant 1 isolate ATV.v1, complete genome			MoxR-type	AAA-ATPase viral structural protein
AE006641.1_1305634_1311637_32_spacer_1307614_38	Phage	KX607102.1		4897	4860	CAS-I-A	Sulfolobus_solfataricus_P2_GCA_00Acidianus two-tailed virus variant 1 isolate ATV.v1, complete genome			unknown	
AE006641.1_1305634_1311637_34_spacer_1307738_39	Phage	HG322870.1		33321	33359	CAS-I-A	Sulfolobus_solfataricus_P2_GCA_00Sulfolobus monocaudavirus SMV1 complete genome			unknown	
AE006641.1_1305634_1311637_35_spacer_1307801_39	Phage	HG322870.1		4469	4431	CAS-I-A	Sulfolobus_solfataricus_P2_GCA_00Sulfolobus monocaudavirus SMV1 complete genome			unknown	
AE006641.1_1305634_1311637_36_spacer_1307864_37	Phage	CP001405		10685	10721	CAS-I-A	Sulfolobus_solfataricus_P2_GCA_00Sulfolobus islandicus Y.N.15.51 plasmid pYN01, complete sequence			transposase,	IS605 OrfB family
AE006641.1_1305634_1311637_37_spacer_1307925_38	Phage	HG322870.1		36946	36909	CAS-I-A	Sulfolobus_solfataricus_P2_GCA_00Sulfolobus monocaudavirus SMV1 complete genome			intergenic	
AE006641.1_1305634_1311637_39_spacer_1308050_40	Phage	HG322870.1		42925	42886	CAS-I-A	Sulfolobus_solfataricus_P2_GCA_00Sulfolobus monocaudavirus SMV1 complete genome			conserved	archaeal protein
AE006641.1_1305634_1311637_40_spacer_1308114_40	Phage	HG322870.1		9376	9337	CAS-I-A	Sulfolobus_solfataricus_P2_GCA_00Sulfolobus monocaudavirus SMV1 complete genome			conserved	archaeal viral protei
AE006641.1_1305634_1311637_9_spacer_1306161_39	Phage	KX607102.1		42367	42329	CAS-I-A	Sulfolobus_solfataricus_P2_GCA_00Acidianus two-tailed virus variant 1 isolate ATV.v1, complete genome			archaeal	transmembrane protein

**Organism**

## Arrays found

ContigId  
 CP003056.1  
 CP003056.1  
 CP003056.1  
 CP003056.1

**Bacillus coagulans 36D1**

Start	Stop	Size	Repeat	Spacers	CRISPR Type
2117433	2121795	4362	GTTTGTATTTTACCTATGAGGAATTGAAAC	65	CAS-I-B
2123872	2124703	831	GTTTGTATTTTACCTATGAGGAATTGAAAC	12	CAS-I-B
2126172	2127007	835	GTTTGTATTTTACCTATGAGGAATTGAAAC	12	CAS-I-B
1096065	1097943	1878	GTTTCAATTCCTCATAGTAAATACTAAC	28	Unidentified

Organism  
 Bacillus\_coagulans\_36D1\_GCA\_000169195.2  
 Bacillus\_coagulans\_36D1\_GCA\_000169195.2  
 Bacillus\_coagulans\_36D1\_GCA\_000169195.2  
 Bacillus\_coagulans\_36D1\_GCA\_000169195.2

## Hits

SpacerID  
 CP003056.1\_1096065\_1097943\_8\_spacer\_1096556\_35  
 CP003056.1\_2117433\_2121795\_14\_spacer\_2118331\_37  
 CP003056.1\_2117433\_2121795\_16\_spacer\_2118466\_39  
 CP003056.1\_2117433\_2121795\_3\_spacer\_2117598\_36  
 CP003056.1\_2117433\_2121795\_36\_spacer\_2119791\_36  
 CP003056.1\_2117433\_2121795\_38\_spacer\_2119925\_35  
 CP003056.1\_2117433\_2121795\_51\_spacer\_2120803\_37  
 CP003056.1\_2126172\_2127007\_1\_spacer\_2126202\_36  
 CP003056.1\_2126172\_2127007\_9\_spacer\_2126738\_37

Protospacer	ProtospacerCt	ProtospacerEnd
Phage	JZDH0100002I	1454
Phage	LBMQ010000I	1105
Phage	DS483513	1432
Phage	LBMQ010000I	517
Phage	JZDH0100002I	2448
Phage	JZDH0100002I	33751
Phage	LBMQ010000I	12639
Phage	LBMQ010000I	11525
Phage	LBMQ010000I	44036

CRISPR Typ	CRISPR Organism	Protospacer Orga	Target ann	Note
1488	Unidentifie	Bacillus_coagulans_36D1_GCA_000169195.	Bacillus coagulans Type II seci	prophage region
1141	CAS-I-B	Bacillus_coagulans_36D1_GCA_000169195.	Bacillus coagulans unknown	prophage region
1470	CAS-I-B	Bacillus_coagulans_36D1_GCA_000169195.	Parvimonas micra unknown	prophage region
552	CAS-I-B	Bacillus_coagulans_36D1_GCA_000169195.	Bacillus coagulans unknown	prophage region
2483	CAS-I-B	Bacillus_coagulans_36D1_GCA_000169195.	Bacillus coagulans unknown	prophage region
33785	CAS-I-B	Bacillus_coagulans_36D1_GCA_000169195.	Bacillus coagulans unknown	prophage region
12675	CAS-I-B	Bacillus_coagulans_36D1_GCA_000169195.	Bacillus coagulans unknown	prophage region
11560	CAS-I-B	Bacillus_coagulans_36D1_GCA_000169195.	Bacillus coagulans unknown	prophage region
44072	CAS-I-B	Bacillus_coagulans_36D1_GCA_000169195.	Bacillus coagulans unknown	prophage region

**Organism**

## Arrays found

ContigId  
 AP014696.1  
 AP014696.1  
 AP014696.1  
 AP014696.1  
 AP014696.1

**Clostridium botulinum strain 111**

Start	Stop	Size	Repeat	Spacers	CRISPR Type
2305105	2306712	1607	ATTTAAATA	17	CAS-I-B
2316191	2316671	480	ATTTAAATA	7	CAS-I-B
2327560	2331536	3976	ATTTAAATA	33	CAS-III-B
2337096	2337798	702	ATTTAAATA	10	Unidentified
2350475	2351031	556	ATTTAAATA	8	Unidentified

Organism  
 Clostridium\_botulinum\_111\_GCA\_000829015.1  
 Clostridium\_botulinum\_111\_GCA\_000829015.1  
 Clostridium\_botulinum\_111\_GCA\_000829015.1  
 Clostridium\_botulinum\_111\_GCA\_000829015.1  
 Clostridium\_botulinum\_111\_GCA\_000829015.1

## Hits

SpacerID  
 AP014696.1\_2305105\_2306712\_13\_spacer\_2306381\_37  
 AP014696.1\_2305105\_2306712\_15\_spacer\_2306514\_36  
 AP014696.1\_2305105\_2306712\_4\_spacer\_2305572\_34  
 AP014696.1\_2305105\_2306712\_5\_spacer\_2305636\_36  
 AP014696.1\_2316191\_2316671\_2\_spacer\_2316278\_36  
 AP014696.1\_2316191\_2316671\_3\_spacer\_2316344\_35  
 AP014696.1\_2316191\_2316671\_4\_spacer\_2316409\_34  
 AP014696.1\_2327560\_2331536\_10\_spacer\_2329090\_34  
 AP014696.1\_2327560\_2331536\_26\_spacer\_2331016\_35  
 AP014696.1\_2327560\_2331536\_27\_spacer\_2331081\_34  
 AP014696.1\_2327560\_2331536\_28\_spacer\_2331145\_34  
 AP014696.1\_2327560\_2331536\_29\_spacer\_2331209\_36  
 AP014696.1\_2327560\_2331536\_30\_spacer\_2331275\_35  
 AP014696.1\_2337096\_2337798\_1\_spacer\_2337126\_34

ProtospacerType	ProtospacerContig	Protospacer	Protospacer	CRISPR Typ	CRISPR Organism
Phage	JXMY01000028	127547	127583	CAS-I-B	Clostridium_botulinum_111_GCA_000829015.1
Phage	LFQ01000066	1413	1448	CAS-I-B	Clostridium_botulinum_111_GCA_000829015.1
Phage	LAGN01000021	36058	36091	CAS-I-B	Clostridium_botulinum_111_GCA_000829015.1
Phage	JH470460	5937	5972	CAS-I-B	Clostridium_botulinum_111_GCA_000829015.1
ORF	LAGK01000703	1765	1800	CAS-I-B	Clostridium_botulinum_111_GCA_000829015.1
Phage	JH470460	37236	37270	CAS-I-B	Clostridium_botulinum_111_GCA_000829015.1
Phage	LFPO01000007	34138	34171	CAS-I-B	Clostridium_botulinum_111_GCA_000829015.1
ORF	JSCF01000042	1388	1421	CAS-III-B	Clostridium_botulinum_111_GCA_000829015.1
Phage	LFOQ01000040	28176	28210	CAS-III-B	Clostridium_botulinum_111_GCA_000829015.1
Phage	LFQ01000043	18707	18740	CAS-III-B	Clostridium_botulinum_111_GCA_000829015.1
Phage	JXSU01000008	103302	103335	CAS-III-B	Clostridium_botulinum_111_GCA_000829015.1
Phage	LAGH01000324	4021	4056	CAS-III-B	Clostridium_botulinum_111_GCA_000829015.1
ORF	LFQD01000008	47485	47519	CAS-III-B	Clostridium_botulinum_111_GCA_000829015.1
Phage	JXSU01000007	2875029	2875062	Unidentifie	Clostridium_botulinum_111_GCA_000829015.1

Protospacer Organism	Target annotat	Note
Clostridium_botulinum_111_GCA_000829015.1	Clostridium botulinum tail protein	prophage region
Clostridium_botulinum_111_GCA_000829015.1	Clostridium botulinum unknown	prophage region
Clostridium_botulinum_111_GCA_000829015.1	Clostridium botulinum unknown	prophage region
Clostridium_botulinum_111_GCA_000829015.1	Clostridium sporogenesTraG/TraD fam	prophage region and plasmid replication
Clostridium_botulinum_111_GCA_000829015.1	Clostridium botulinum unknown	
Clostridium_botulinum_111_GCA_000829015.1	Clostridium sporogenesunknown	prophage region
Clostridium_botulinum_111_GCA_000829015.1	Clostridium botulinum unknown	prophage region
Clostridium_botulinum_111_GCA_000829015.1	Clostridium botulinum unknown	prophage region
Clostridium_botulinum_111_GCA_000829015.1	Clostridium botulinum unknown	prophage region
Clostridium_botulinum_111_GCA_000829015.1	Clostridium sporogenesunknown	prophage region
Clostridium_botulinum_111_GCA_000829015.1	Clostridium botulinum unknown	
Clostridium_botulinum_111_GCA_000829015.1	Clostridium botulinum nucleotide pyroprophage region	