

table S1 Summary of contigs containing template repeat/ variable repeat pairs and RTs. Variable region sequences are in Fig. S1.																											
Contig name	Subject	Length	Family	Variable region	ORF Coordinates	RT Coordinates	Repeated Regions	BLASTp/NR (orf containing variable region)						Phyre2 structure prediction over variable region				Phyre2 structure prediction over non-variable region				GC-only near VR					
								Name	E-value	%Identity	%Length	Name	Model (P98)	Confidence	Coverage	Name	Model (P98)	Confidence	Coverage	C-type lectin	lg		LRR	Helix in VR	Helix in ORF		
1013_scaffold1563	1013	42703	N	22583-22683	16207-22752	23414-24001	22583-22683 - 22869-22969	putative phage head-tail adaptor [Capnocytophaga sp. oral taxon 329 str. F0087]	332879328	36%	91%	0	None	N/A	N/A	N/A	LRR	30ja	95%	12%		X	No	8607-8632	No	No	
1013_scaffold1713	1013	13658	N	8562-8788	9807-8650	7231-6851	7998-8136 - 14041-14129 - 13625	None	N/A	N/A	N/A	N/A	None	N/A	N/A	N/A	None	N/A	N/A	N/A						No	No
1013_scaffold1877	1013	19577	M/P	14674-14762	15694-14669	13625	14674-14762	Chain A, Major Tropism Determinant U1 Variant	78101014	29%	99%	6.00E-20	MTD	2iou	100%	87%	None	N/A	N/A	N/A	X			No		No	
1013_scaffold24	1013	47751	M	40232-40356	41594-40227	39050-38541	39459-39555 - 40230-40345	None	N/A	N/A	0%	0	None	N/A	N/A	N/A	None	N/A	N/A	N/A			40199-40224	No	No		
1013_scaffold3125	1013	94489	S/M	47601-47741	49656-47602	45627-46016	46739-46866 - 47606-47735	None	N/A	N/A	N/A	N/A	None	N/A	N/A	N/A	None	N/A	N/A	N/A						No	No
1013_scaffold47	1013	34299	S/P	18094-18194	19098-18088	17075-16464	17450-17550 - 18094-18194	Chain A, Major Tropism Determinant I1 Variant	78101013	29%	100%	8.00E-18	MTD	2iou	100%	88%	None	N/A	N/A	N/A	X			No		No	
1013_scaffold7838	1013	6526	N	1720-1824	2397-544	3244-2645	3782-3882 - 1728-1828	probable transcriptional regulator-related protein [Clostridium botulinum Bf]	168183960	38%	37%	3.00E-29	cadherin-1	3q2v	98%	74%	None	N/A	N/A	N/A		X		No		No	
2011_scaffold1080	2011	5342	N	3661-3791	1782-3782	4737-5333	3664-3779	None	N/A	N/A	N/A	N/A	None	N/A	N/A	N/A	None	N/A	N/A	N/A			3793-3818	No	No		
2011_scaffold13	2011	55843	S/M	3365-3503	4519-3338	2091-2657	3148-3248 - 5598-5725	None	N/A	N/A	N/A	N/A	None	N/A	N/A	N/A	None	N/A	N/A	N/A			3507-3531	Yes	No		
2011_scaffold152	2011	9198	P	4927-5120	4094-5104	6103-6714	4091-5098	Chain A, Major Tropism Determinant U1 Variant	78101014	30%	100%	1.00E-19	MTD	2iou	100%	97%	None	N/A	N/A	N/A	X		No		4991-5002		
2011_scaffold33	2011	43586	M	39471-39601	40577-39477	38443-37858	38849-38976 - 39482-39609	Chain A, Major Tropism Determinant U1 Variant	78101014	31%	92%	4.00E-20	MTD	2iou	100%	90%	None	N/A	N/A	N/A	X		39618-39643	Yes	No		
2016_scaffold57	2016	45505	S/M	7111-7261	8140-7073	6002-5340	6556-6578 - 7175-7197	ORF35 [Vibrio phage VHML]	27311203	32%	95%	4.00E-36	MTD	2iou	100%	90%	None	N/A	N/A	N/A	X		7019-7044	No	7161-7170		
2019_scaffold132	2019	37433	S	22045-22122	23251-22016	21129-20662	21703-21780 - 22045-22122	None	N/A	N/A	N/A	N/A	MTD	2iou	99%	55%	None	N/A	N/A	N/A	X		21977-22001	No	No		
2019_scaffold260	2019	8655	N	5422-5545	6937-5135	4547-3840	4920-5040 - 5148-5268	phage-related hypothetical protein [Rhodospirillum centenum SW]	209964416	25%	56%	2.00E-13	MTD	2iou	99%	45%	None	N/A	N/A	N/A	X		No		No		
2020_scaffold1264	2020	27831	N	4681-4794	8920-4610	3667-3248	4306-4406 - 4676-4776	putative phage head-tail adaptor [Capnocytophaga sp. oral taxon 329 str. F0087]	332879328	37%	99%	0	Clec	1y4j	90%	23%	LRR	30ja	97%	16%	X	X	No		4660-4670		
2200_scaffold1335	2200	38128	S/M	11391-11503	9722-11524	12112-12819	11625-11745 - 11397-11517	phage-related hypothetical protein [Rhodospirillum centenum SW]	209964451	28%	55%	4.00E-14	MTD	2iou	99%	45%	None	N/A	N/A	N/A	X		No		No		
2200_scaffold2278	2200	36736	S/M	34681-34807	34134-35988	33449-33886	32701-32801 - 34226-34826	probable transcriptional regulator-related protein [Clostridium botulinum Bf]	168183960	37%	37%	9.00E-33	ep-cadherin	1q5c	98%	24%	None	N/A	N/A	N/A		X	34701-34725	Yes	No		
2200_scaffold2352	2200	13046	M	9381-9735	10503-9463	8395-7733	8972-8972 - 9526-9626	ORF35 [Vibrio phage VHML]	27311203	32%	96%	4.00E-33	MTD	2iou	100%	90%	None	N/A	N/A	N/A	X		9409-9434	No	No		
2203_scaffold802	2203	55338	S/M	41681-41781	36628-41778	43266-43658	41651-41770 - 42437-42556	putative phage head-tail adaptor [Capnocytophaga sp. oral taxon 329 str. F0087]	332879328	25%	68%	2.00E-63	None	N/A	N/A	N/A	LRR	30ja	97%	15%		X	41802-41827	No	No		
2204_scaffold14	2204	27737	N	24988-25108	25707-24340	26558-25998	27017-27137 - 24988-25108	None	N/A	N/A	N/A	N/A	None	N/A	N/A	N/A	cellulose-binding protein	2yhg	96%	21%		X	25097-25121	Yes	No		
2204_scaffold211	2204	15588	S	3262-3346	2021-3379	4204-4671	3464-3548 - 3262-3346	None	N/A	N/A	N/A	N/A	Clec	1z70	97%	57%	None	N/A	N/A	N/A	X		No		No		
2204_scaffold812	2204	48707	S/M	26241-26391	27332-26241	25178-24513	25676-25790 - 26259-26373	ORF35 [Vibrio phage VHML]	27311203	33%	95%	2.00E-29	MTD	2iou	100%	91%	None	N/A	N/A	N/A	X		26193-26217	No	No		
2207_scaffold2979	2207	5560	N	1731-1841	841-1890	2722-3183	2108-2258 - 1742-1892	None	N/A	N/A	N/A	N/A	MTD	2iou	98%	72%	None	N/A	N/A	N/A	X		No		No		
2207_scaffold314	2207	8886	N	4081-4231	2597-4276	5219-5638	4471-4529 - 4101-4159	putative phage head-tail adaptor [Capnocytophaga sp. oral taxon 329 str. F0087]	332879328	48%	98%	2.00E-135	MTD	2iou	97%	43%	None	N/A	N/A	N/A	X		No		4220-4230		
2209_scaffold1451	2209	19156	S/M	1945-1988	2574-715	3412-2813	3993-4036 - 1945-1988	probable transcriptional regulator-related protein [Clostridium botulinum Bf]	168183960	39%	37%	3.00E-33	cadherin-1	3q2v	100%	77%	None	N/A	N/A	N/A		X	1989-2013	Yes	No		
2209_scaffold2350	2209	33654	S	25528-25628	25636-25046	24047-24652	23576-23676 - 25528-25628	None	N/A	N/A	N/A	N/A	None	N/A	N/A	N/A	None	N/A	N/A	N/A			No		No		
2209_scaffold441	2209	29952	N	6471-6591	5839-7545	5166-5582	4412-4512 - (6484:6584 AND 16553:16653)	gp427 [Bacillus phage G]	34749644	42%	28%	2.00E-11	receptor-type tyrosine-protein phosphatase mu	2v5y	100%	65%	None	N/A	N/A	N/A		X	No		No		
2209_scaffold64	2209	42319	N	21121-21291; 34511-34631	33896-34834	19555-19148	20259-20359 - (20919:21019 AND 34,194:34,294)	None	N/A	N/A	N/A	N/A	None	N/A	N/A	N/A	None	N/A	N/A	N/A			34141-34165	Yes	No		
2210_scaffold709	2210	36303	M	11132-11252	11900-10329	12744-12145	13118-13238 - 11132-11252	TII protein [Clostridium lentocellum DSM 5427]	326791598	27%	50%	3.00E-10	endoglucanase	2c2c	100%	41%	None	N/A	N/A	N/A		X	No		11181-11190		