Supplementary Material for

Diversity and abundance of single-stranded DNA viruses in human faeces

Min-Soo Kim, Eun-Jin Park, Seong Woon Roh and Jin-Woo Bae*

*To whom correspondence should be addressed. E-mail: baejw@khu.ac.kr

This PDF file includes:

Figs. S1 to S3

Tables S1 to S4

Individual	Sample ID	Birth Country	Age	Sex	Sample amount
Individual 1	F-A		28	Male	12.4g
Individual 2	F-B		29	Male	15.2g
Individual 3	F-C	South Korea	28	Male	14.4g
Individual 4	F-D		23	Female	14.1g
Individual 5	F-E		27	Female	12.4g

Table S1. Features of participants and samples.

							No.	of sequences	(%)							
-			SEED_nr					CAMERA_v	r				CAMERA_n	r		
	BLASTx ($E < 10^{-5}$)						BI	LASTx ($E < 1$	0 ⁻⁵)		BLASTx (<i>E</i> <10 ⁻³)					
-	F-A	F-B	F-C	F-D	F-E	F-A	F-B	F-C	F-D	F-E	F-A	F-B	F-C	F-D	F-E	
Raw reads	113,054	109,569	68,391	115,121	98,511	113,054	109,569	68,391	115,121	98,511	113,054	109,569	68,391	115,121	98,511	
Quality-filtered	100,358	98,601	61,641	107,131	86,483	100,358	98,601	61,641	107,131	86,483	100,358	98,601	61,641	107,131	86,483	
Non-redundant	90,273 (100%)	88,810 (100%)	53,342 (100%)	95,902 (100%)	71,806 (100%)	90,273 (100%)	88,810 (100%)	53,342 (100%)	95,902 (100%)	71,806 (100%)	90,273 (100%)	88,810 (100%)	53,342 (100%)	95,902 (100%)	71,806 (100%)	
Unknown	80,523 (89.2%)	78,989 (88.94%)	47,246 (88.57%)	86,777 (90.49%)	62,534 (87.09%)	81,738 (90.55%)	83,190 (93.67%)	49,350 (92.5%)	81,694 (85.18%)	63,447 (88.36%)	78,234 (86.66%)	71,838 (80.89%)	43,535 (81.61%)	69,779 (72.76%)	54,422 (75.79%)	
Known	9,750 (10.8%)	9,821 11.06%)	6,096 (11.43%)	9,125 (9.51%)	9,272 (12.91%)	8,535 (9.45%)	5620 (6.33%)	3,992 (7.5%)	14,208 (14.82%)	8,359 (11.64%)	12,039 (13.34%)	16,972 (19.11%)	9,807 (18.39%)	26,123 (27.24%)	17,384 (24.21%)	
Virus	6,091 (6.75%)	2,125 (2.39%)	1,902 (3.57%)	2,976 (3.1%)	7,000 (9.75%)	8,467 (9.38%)	5,620 (6.33%)	3,952 (7.4%)	14,155 (14.76%)	8,348 (11.63%)	6,186 (6.85%)	2,381 (2.68%)	1,274 (2.39%)	10,255 (10.69%)	4,049 (5.64%)	
Bacteria	3,599 (3.99%)	6,709 (7.55%)	4,103 (7.69%)	6,073 (6.33%)	2,180 (3.04%)	8 (0.01%)	0	35 (0.07%)	28 (0.03%)	0	5,661 (6.27%)	13,944 (15.7%)	8,263 (15.49%)	15,758 (16.43%)	13,198 (18.38%)	
Archaea	13 (0.01%)	711 (0.8%)	13 (0.02%)	10 (0.01%)	19 (0.03%)	0	0	0	0	0	0	316 (0.36%)	0	0	0	
Eukarya	47 (0.05%)	276 (0.31%)	78 (0.15%)	64 (0.07%)	73 (0.1%)	60 (0.07%)	0	5 (0.01%)	25 (0.03%)	11 (0.02%)	157 (0.14%)	133 (0.15%)	153 (0.29%)	86 (0.09%)	106 (0.15%)	
Other	0	0	0	2 (0.002%)	0	0	0	0	0	0	35 (0.04%)	198 (0.22%)	117 (0.22%)	24 (0.03%)	31 (0.04%)	

Table S2. The BLAST profile of five viromes comparing with three public databases, the SEED nr, CAMERA nr and CAMERA viral protein databases.

 Table S3. The profiles of viral families of five viromes comparing with three public databases, the SEED nr, CAMERA nr and CAMERA viral protein databases.

				SFFD n	r		CAMERA v						CAMERA nr						
			ы	SEED_I	10 ⁻⁵			$\mathbf{D}_{\mathbf{A}} \otimes \mathbf{T}_{\mathbf{a}} = (\mathbf{a} < 10^{-5})$						$\mathbf{D} \mathbf{I} \mathbf{A} \mathbf{C} \mathbf{T} = (\mathbf{D} \mathbf{C} 10^{-3})$					
			BL	481x (E <	<10)			E D	ASTX (E	<10')			BLA	481x (E ·	<10')				
		F-A	F-B	F-C	F-D	F-E	F-A	F-B	F-C	F-D	F-E	F-A	F-B	F-C	F-D	F-E			
Viral sequence	es	6,090	2,125	1,902	2,976	7,000	8,467	5,620	3,952	14,155	8,348	6,186	2,381	1,274	10,255	4,049			
dsDNA	Муо-	51	240	162	210	76	362	877	328	132	134	19	165	50	22	47			
	Podo-	4,252	1,144	323	778	801	4,928	1,834	507	695	1,455	4,772	1,948	370	735	1,265			
	Sipho-	121	224	289	278	47	702	1,291	1,043	10,381	203	42	41	413	8,809	15			
	Unclassified Caudovirales	3	20	12	2	74	138	252	169	264	131	0	0	15	92	60			
	Asfar-	1	0	1	1	0	6	0	5	0	0	7	0	6	0	0			
	Irido-	0	11	1	0	0	27	44	0	0	0	13	6	0	2	0			
	Herpes-	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0			
	Adeno-	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0			
	Mimi-	4	29	5	5	2	14	28	29	20	0	3	3	6	7	0			
	Phycodna-	11	22	2	0	3	23	24	6	0	6	0	5	0	0	0			
	Pox-	5	55	13	0	0	8	19	20	0	0	0	0	0	0	0			
	Tecti-	0	0	2	13	0	16	0	0	10	9	9	0	0	4	0			
	Bicauda-	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0			
	Unclassified viruses	0	2	0	0	0	102	89	90	5	26	2	4	4	2	0			
ssDNA	Ino-	0	0	0	0	0	0	12	0	0	0	0	0	0	0	0			
	Anello-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2			
	Germini-	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0			
	Micro-	1,638	377	1,087	1,655	5,993	1,800	420	1,182	2,467	6,211	1,312	207	408	576	2,660			
	Unclassified-	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0			
Unclassified	Unclassified phages	0	1	3	26	1	339	723	566	181	173	5	2	2	0	0			
RNA	Picorna-	4	0	2	4	3	0	0	0	0	0	2	0	0	2	0			

	Sample	Richness ^a	Shannon-Wiener index (nats)
	F-A	34	3.04
	F-B	24	2.79
Viral assemblage	F-C	26	2.84
	F-D	18	2.58
	F-E	401	4.19
	F-A	1636 / 2510	4.64
	F-B	1214 / 1632	4.25
Bacterial community	F-C	1607 / 2197	4.55
	F-D	914 / 1247	4.67
	F-E	1492 / 1953	4.59

Table S4. Comparison of viral and bacterial community structures and diversity.

^a In bacterial richness, Chao1 and Ace estimators are shown in order.



Fig. S1. Functional gene profiles of five viromes comparing to the SEED database. Both the sequence reads and large contigs (>500 bp) of five viromes were compared in BLAST searches (BLASTx, E-value < 10⁻²). The functional categories including higher than 1% of the identified sequences were shown.

	10			10	50		70 00		
	10	20	30	40	50	60	70 . 80	90 100	
EnterobacteriaphageWA13_YP_512416/1-231	ID IMGLOAAYAQLI	TEGERDYFMTE	RDIVNEFGG	STSYDA-DN	RPLLVMHSEFWA	SGYDVDGTD	Q	SSLGOFSGRVQQTFKHSVPR	FFCPEH
EnterobacteriaphageID18_YP_512796/1-230	ID IMGLOAAYAKLI	TEGERDYFMTR	RDIMKDFGG	HTSYDG-DN	RPLLLMRSEFWA	SGYDVDGTD	Q	TSLGOFSGRVQQTFNHKVPR	FYVPEH
EnterobacteriaphageG4 NP 040678/1-231	ID IMGLOAAYAKLI	TEGERDYFMTR	RDIMKEFGG	HTSYDG - DN	PLLLMRSEFWA	SGYDVDGTD	· · · · · · · · · · · · · · · · · · ·	SSLGOFSGRVQQTFNHKVPR	FYVPEH
Enterobacteriaphage/D2_YP_512376/1-231	ID IMGLOAAYAKLI	TODERDYEMTE	RDIMKDEGG	HTSYDG - DN	PPLLLMRSEFW	SGYDVDGT0	G	SSLGOF SGRVQQTENHK VPR	FYVPEH
EnterobacteriaphagephiX174 NP_040711/1-231	ID IMGLOAAYANLI	TDOERDYFMOR	HOVISSEGG	KTSYDA-DN	RPLLVMRSNLW/	SGYDVDGTD	Q	TSLOOFSGRVQQTYKHS VPR	FFVPEH
Enterobacteriaphage alpha3 NP 039597/1-235	ID IMOL QAAYAQL	TEGERTYEMOR	ROVISSEGG	STSYDA-DN	PLL VMHTDEWA	SGYDVDGTD		SSL GOFSGRVOOTEKHS VPR	FEVPEN
EnternhacterianhaneSt.1 VP 002095212/1-235	ID IMOLOAAYAOL	TEDERTYEMOR	POVISSEGG	STSYDA DN	PRI I VMHTDEWA	SAVANARTO		SSI GOESGRVOOTEKHS - VPR	FEVPEH
April 20185 E-2/1-225	AT INCLERALAL OF	EVEISADARSD	IFTIMOLEH	VDBSDADIO	DEVI DOGVINI	DIGEVIDISA	TO E	POCHMACPECIERSN. OPT	VDAEEH
0074000475 E 4/1-255	UT INDI PRATALO	EVELSADAGOD	VENINGHEN	VTDEDADLD	PEVI CCCVST	NITEIPOTRA	F E	POONT ADVOCATORSN	FEREN
ADDIAD00000 E AIA 244		EVELOADACOD		NTROPARI OF	DEVI COOVETS	NITE POTOA	0 5		E E A C B C
CORREGIOUSS F-W1-211	THE PLAN ALL	ETELSA AGO		VIPADARLU	PETLOGGVSTV	R TE POTOA	La	SPUCHLAGAGE BIGRON ROO	FF AC
spiropiasma_pnage4/1-241	BISDLELATING	INKEMDAHGG H	VEFILNHEG	VHIADAHLU	Rath Loghaval	LVUSVPUTAS	VEKM	PUGNLAAPSEIMIGNNY-LVN	KELER
contig00760_F-C/1-240	ATTNDLHQATALQ	HIFESDAHNGIN	REFLIGING	VISPOSELQ	TPEYIGGORIAI	NVNQVVQTSQT	1 1	QALGNTAAMSLTTCSKQ MAU	YAATEY
contig00491_F-DV1-239	FSVNELHLAFAYO	FLESLAPSGSP	ELLLGLFG	VRSPDARLO	PEYLOGNRVPI	NVSEVINSAQ	8EQ	DELGOLGAKSSTSDVNHDEV	KSFIEH
contig00850_F-A/1-239	ATINGIPOAFOVO	WFEELARGGSR	REGIVELER	TRISDKTVO	I PEYLOGDR I MI	NMSQVVQTSG	TV	SPOGNVSAMSVIGFOKS AFT	KSFEEH
contig00848_F-A/1-239	ATINQLROAFQVQ	YFEELARGGSR	REQIVELER	RISOKTVQ	I PEYLOGDR I MI	INMNOVIQISE	TET	SPOGNVAAMSVIGFOKS AFT	KSFEEH
contig00544_F-E/1-239	ATINGLEGAFOVO	YFEELARGGSR	REQIYSLER	TRISDKTVO	I PEYLGGDR I MI	NMNQVIQTSG	TET	SPOGNVAAMSVIGFOKS AFT	KSFEEH
contig00061_F-D/1-238	ITINEL ROAFAVO	YYEALARGGSR	REQVEALES	VSISDKTVO	VPEYLGGGRYHV	NINQIVOTSGQC	TETD	TPIGETGAMSVTPINES SFT	KSFEEH
contig00543_F-E/1-238	TTINGLEGAFAVO	YYEALARGESRY	REQVEALES	VEISOKTVO	VPEYLOGGRYHN	NVNOI VOT SCOC	2 T TE	TPICETCAMSVTPINES SFT	KSLEEH
Bdellovibrio_phage_phiMH2K/1-239	AT INQL ROAMMMQ	SLLELDARGGIR	VEILKSHEN	VISLDFRLQ	RPEYLSGGTIDL	QQNPVPQTSS	STTD	SPQ <mark>GNLAAF</mark> STASEF <mark>G</mark> NKIGFS	KSEVEH
ChlamydiaphageChp1/1-243	VTINSLENAITLO	WFEKSARYGSR	VESVOGHEG	VHLGDYRAQ	RP I YLGGSKS YV	SVNPVVQNSS	TDSV	SPOGNLSAYALSTDTKHLFT	KSEVEN
ChlamydophilapsittaciphageChp1/1-243	VTINSLENAITLO	WFEKSARYGSR	VESVOGHEG	VHLGDYRAQ	RP I YLGGSKSYV	SVNPVVQNSS	TDSV	SPOGNLSAYALSTDTKHLFT	KSEVEN
contig00905_F-B/1-236	ATINGLROAFQIQ	LLEKDARGGTRY	REVLREHEG	VISPDSRMQ	I PEYLGGYRLPI	NVSQVIQTSS	TDST	SPLGNTAAL SVTTMNKP MFT	KSFTEH
contig00846 F-A/1-241	ATINGLELAFOIO	FYEROARGOSRY	TEVIRSFFG	VTSPDARLO	RPEYLGGNRVPI	NINOVIOOSG	TGSESSS	TPOSTVVGMSQTTDTNSDFT	KSFTEH
contia00287 F-E/1-239	ATINGLELAFOIO	(LYEKDARGGTR)	TEILKTHEG	VISPDSRLQ		NINOVVONSA	TVEGE	TPLONVACYSVTSDTHS DER	OSFTEH
contio00847 F-A/1-239	ATINGLELAFOIO	LYERDARGGTR	TEILKSHFG	VTSPDARLO	RPEYLGGNRIPI	TINOVVONSG	T MSG E	TPOGTTTAYSLTTDVHQEFT	KSEVEH
contio00907 F-B/1-252	ATINGLELAFOIO	LYERDARGGTR	LEILKSHEG	VISPDARLO	RETLOGNEIPI	NINOVVOSSS	TDAS	TPOGNTAAYSL TSDNHS DFT	KSEVEH
contio00288 F-C/1-239	ATINGLELAFOIO	LYERDARGGTR	TEILKSHEG	VISPDARLO		NINOVVOSSS	TDAS	TPOGNTAAYSL TSDNHS - DFT	KSEVEH
contin00401_E_0/1-239	AT INFL PLAFOLO	(LYERDARGGTR)	I FULKSHEG	VISPOARLO	PEVICOEPIPI	NIDOVIOTSC	TME	TROGNICANSI TONOGS YEK	HSEVEH
Contin513 Anternin (ska/1.230	AT INCL POAFOLO	MYEPDADCCTP	TELIDONEC	VISPDADIO	DEVI COORSYL	OVNSVAOTSP.	99. 00	TPLANI AAFOTCAVIEN OFA	VSEVEN
COS 40500/4.235	STINAL PEAFOLO			VISPOARLO	DEVI COOLTEN	SMODI DOTOS	TO	SPOONL SAMOTVOVOGH CES	DEFUEL
aprila00750 E.C(6-225		L VEDDADCOTD	TELLBOUED	VICODADIO	DEVI COSEAD	SUNDVOOTOA	TN D T	TROOMLAAFOV/VEDEEH	
CONN900739_F=071=233	I BINDL BAT GIG	LYEDDAROOT		V BODALLO	PETEODO AR	STREVOOT SA	The second secon		
Contriguo273_F-E/7-235	ISTNUL PUAPUTO	LYERDARGGIR	I EILESHEG	VISPUARLU	PETLOGSSARI	SINPVUUISA	- NE	PUGNLAAPGVVSDSFHGFS	
ChiamydiaphageChp2/1-243	VIINSLREAFULU	LYERDARGGIR	I ET INSHEN	VUSPDARLU	RAEYLGGSSIPV	NISPIPUISS	U	SPUGNLAAYGTATGSKR VET	NOT THE
Gniamydiaphage3/1-243	VIINSLREAFULU	IL YERDARGGIR	TETTRSHEN	IVUSPDARLU	RAEYLGGSSIPV	NISPIPUISS	10	SPUGNLAAYGTATGSKR VET	KSEIEH
Ghiamydiaphage4/1-245	VTINSLREAFOLO	KLYERDARGGTR		IVQSPDARLQ	RAEYLGGSSTPV	NISPIPOTSS	TDST	SPOGNLAAYGTAIGSKR VFT	KSETEH
ChlamydiaphageCPAR39/1-245	VTINSLREAFOLO	KLYERDARGGTRN	IEIIRSHEN	IVOSPDARLO	RAEYLGGSSTPV	NISPIPATSS	T DST	SPOGNLAAYGTAIGSKR VFI	KSFTEH
GOS_8911/1-235	ASINSLREAFOLO	MLEKDARGGTR	IEIIKSHFG	VTSPDARLQ	RPEYLGGDSQR I	NISQIPOTSS	T D <u>A</u> T	TPOGNLAAYGYLOGTER GFV	KSFTEH
D4_Highborne_Cay/1-227	ATINGLEOSFOID	LYERDARGGIR	TEIIRSHFG	VISPDARLO	RPEYLGGGS TPI	INVNP LAQTGE	ST	TPOGNLAAMGTAYMDGH GFT	KSFTEH
A3_Highborne_Cay/1-234	ATINGLEGSEGIG	LYERDARGGIR	TEIIRSHFG	VISPDARLO	RPEYLGGGS TPI	NVNP I ADTGE	S G T	TPOGNLAAMGTAYMDGH GFT	KSFTEH
B4_Highborne_Cay/1-234	ATINGLEQSEQIQ	LYERDARGGTRY	TEIIRSHFG	VTSPDARLO	PEYLOGGS TP I	NVNP I AQTGE	S G T	TPOGNLAAMGTAYMDGHGFT	KSFTEH
A1_Highborne_Gay/1-234	ATINGLROSFOIO	LYERDARGGTRY	TEIIRSHFG	VISPDARLO	RPEYLGGGS TP I	NVNP I AQTGE	ST	TPOGNLAAMGTAYMDGH GFT	KSFTEH
B3_Highborne_Cay/1-234	ATINGLEOSFOIO	LYERDARGGTRY	TEIIRSHFG	VISPDARLO	PEYLOGGS TPI	NVNP I AQTGE	S G T	TPOGNLAAMGTAYMDGH GFT	KSETEH
B1_Highborne_Cay/1-233	ATINGLEOSFOIO	LYERDARGGTRY	TEIIRSHFG	VISPDARLO	RPEYLOGGSTPI	NVNP I AQTGE	5T	TPOGNLAAMGTAYMDGH GFT	KSFTEH
A2_Highborne_Cay/1-234	ATINGLEOSFOIO	LYERDARGGTR	TEIIRSHFG	VISPDARLO	PEYLOGGSTPI	NVNP I AQTGE	5 G T	TPOGNLAAMGTAYMDGH GFT	KSFTEH
A4_Highborne_Cay/1-234	ATINGLEOSFOIO	LYERDARGGTR	TEIIRSHFG	VISPDARLO	RPEYLOGGSTPI	NVNP I AQTGE	S G T	TPOGNLAAMGTAYMDGH GFT	KSFTEH
Contig516 Antarctic Jake/1-241	AT INOL POSFOIO	(LLERDARGGTR)	TEIIRSHFG	VVSPDARLO	RPEYIGGGTTPI	IN INPIAOTSA	TGETGAT	TPLGTLASMGTALSHNHGFT	OSETEH
Contig515 Antarotic lake/1-241	ATINOL POSFOIO	LLERDARGGTR	TEIVESHEG	VISPDARLO	RPEYIGGGTTSI	IN INPLACTSG	TGQTGQT	TPLGTLASMGTALAHNH GFT	OSFTEH
GOS 7913/1-241	ATINGLEOSEDIO	LLERDARGGIR	TEILRSHEG	VASPDARLO	RPEYLSGGS TPI	IN LOPVADTSG	TGASGQT	TPIGNLAAYGTYLAQSH GFS	OSEVEN
GOS 11182/1-241	ATINGLRESFOID	LLERDARGGTR	TELIPSHEG	VISPDARLO	PEYLOGGSTPV	NISPIAOTSA	TASGATA	TPLONLAAMGTALAMGH GFT	OSEVEN
GOS 7850/1-241	AT INCLRESEATO	LI ERDARGGTRY	TELLESHER	VVSPDAPL OF	PEVI GGGSAPI	VINPVADISA	ISASGIN	TPL ONLAAVGTAMASCH GET	OSETER
GOS 6000/1-236	AT INCLRESEATO	IL FROAPGGTP	TELLESHEG	VVSPDARL OF	PETLOGOSAPI	VINEVADISA	TSASGT N	TPI GNI AAVGTAMASGH GEVI	OSETEN
GOS 10096/1-241	AT INCL PESEDIC	LIEPDARCCTR	TELLESHEG	VVSPDADIO	PREVI GOOSTRY	SINDIAOTSA	TAAGAG	TPI CNI AAMOTALAMOH CET	OBETEN
GOS 10508/1-241	AT INCL PESEOLO	FERNARCOTE	TELIDSHED	VVSPDADIO	PREVI COOSVPL	SINDVACTSA	TASCAC	TPI CNI AAMOTAL AMOH OFT	OSETEN
GOS 10000/1-241	AT INCLESSOUCH	PLIEPDAPCCTP	TELIPSHEC	VEDBADIO	PREVI COCOTEN	SINDIAOTSA	TASCAC	TPL CHI AAMOTAL AMOH CET	OBETEN
005 100011-241	ATINOL PESEOLO	FRADOCTO	TELIDONEC	VVEDDADIO	PREVI COCOVPL	SINDVACTOA	TAADAD	TOLONI AAMOTALAMOH OFT	OBETEN
202 0/25// 24/	ATING	LLEDDADCOTD	TELIDOUEO	V POPAPIO	DOENL COOD VOL	CINDVACTOA	TOCOAD	TOLONI AAMOTAL AMOU	
000_000000000	AT INCLUSE OF OTO	LLECDAROOT	THURSDAR		DENLOCOD VP	STREVACTOA.	33040	TOL ONLY AND TAL AMORT OF	
000_76004.004	AT THE COLOR	LLERDAHOOTH	TELLEOUCE	V BEDALLU	DOEN COOD VOL	STREVACTOA.		ELGNEAANG ALANGH - GET	
GUS_75797-234	ATTNULHEAFUVU	LLERUNHOGIN	1 III AAAA	A A B P D A P L U	PETLOGGEVPV	STREVAUTSA	AAGAG	TPL GREAAMGTALAMGT GFT	
Prevotella_bergensis_DSM_17301_(ZP_00006/71)/1-248	SVSSLBAAFAVD	LUSVIMRAGKT	UBUMHAHYC	WEI PUSHDG	TLUGFUSD	avsbyruise-	THOREONY ON A	LORVACKUIGSCROR IV	FUAKEH
Prevotella_buccalls_ATCC_35310_(ZP_06286058)/1-250	LIVQSINNAFALD	LMRITORAGKT	AEQUAHEG	FEVSECROG	TIGGEDSNI	QVODVTQMSG	TASPEQOV-SIKHO	GTLORVIGKAQGSGSGHIEI	FUAHEH
Prevotella_sp_oral_taxion_317_str_F0108_(ZP_06421583)/1-25	O LNVSATRSAFALD	LLSISMRAGKT	ABQIEAHFG	VEEGEDGE	UVYYLGGFDSN	QVGDVTQTSG	TNENVSEVGNAKLA	BYLGKIITGKGIGSGYGE IQ	FOAKEP
Bacteroides_plebeius_DSM_17135/1+259	VSILALERAEAAQ	WKEVALASEED	PSQIEAHWG	QSVNKAYSDI	MCOWLOSINIDL	SINEVVNNNI	1 <u>6</u>	ENAADTAGKGTMSGNGS IN	FNVGGQ
Bacteroides_eggerthii_DSM_20697/1-261	LSILVLRQAEFLQ	WKEITQSGNKD	KDOLEKHWG	VSVGDGF SEI	LCTYLCCVSSSI	DINEVININI	I G	SAAADIAGKGVGVANGE IN	FNENGR
Bacteroides/1-254	FTVLALRQAEFLQ	WKEITOSONKO	KDOTEKHWN	VSVGEAYSE	MSLYLGGTTASL	DINEVVNNNI	👖 G	SNAADIAGKGVVVGNGR 15	FDAGER
contig00744_F-B/1-253	FDILSFRIAEATO	WKEVTQCAKQG	KEQLEAHEN	KLSEALSD	HORYIGGTSSGI	TISEVLNTNL	E	SGDAN KGKGVGGSFGFET	FEINER
contig00335_F-D/1-257	FSILOLEMAEAVQ	YREVSOFADOD/	RGOIMAHEG	VSL SPVL SD	KCMYLGGSSSNI	DLSEVVNTNI	🚺	DNIAEIAGKGVGTGQGSFS	GNEDTY
contig00377_F-D/1-257	FSILOLEMAEAVO	REVSQFADQD	RGQIMAHEG	VSLSPVLSD	KCVYLGGSSSNI	DLSEVVNTNI	I G	DNVAEIAGKGVGTGQGSFS	GOFDEY
contig00845_F-A/1-253	FSVLOLRAAEALO	WKELAQANGON	AAOVKAHEG	VSTNPMQSH	STRICGEDGS	DISAVENTNL	T	ADEAI IRGKGLGGORIND - PSD	FICNEH
contig00489_F-D/1-255	LOVLALERGEALO	FREISLCTPLN	RSQIEAHEG	VDVGSELSG	MSTYIGGEASSL	DISEVVNINI	IE	SNEAL IAGKGIGTGQGN EE	FYAKDW

	120	130	140	15	50	160	170	180	100	200	210	220
Enternhaderianhadel/VA13_VD_512416/1-221	OTHET AL VEEDDT	. ENEMAVIO	CHEOL TV	DI ACOR	AL MANE DE	DEVINAS	FERS	100	SAVEVI	ARCON. N	MODOR	AL DUNY
Enterobacteriaphage/D18_VP_512796/1-230	GVINTL SVTREPPT	- EMEMHYLV	GRESLTY	DIACOP	ALMANLPE	RELEVED	LEHS A	1 8	·····KEKI	AEGOW Y	TOPDRY	AFPYNA
EnterobacteriaphageG4_NP_040678/1-231	RVINTLAVTREPPT	- EMEMHYLV	GKENLTY	DIACOP	ALMANLPP	REVELKE	FFHS	PD	SAKEKI	AEGOW Y	TOPDRY	AFPYNA
EnterobacteriaphageID2_YP_512376/1-231	- GVIMTLAVTREPPTI	H - EMEMHYLV	GKESLTY	DLACOP	ALMANLPP	PREVELKE	FFHS	PD	SAKFKI	AEGOW Y	TOPDRY	AFPYNA
EnterobacteriaphagephiX174_NP_040711/1-231	- GTMFTLALVREPPT	A-TKEIQYLN	AKGAL TY	DIAGOP	VLYGNLPF	PRE I SM <mark>KD</mark>	VFRS	DS	SKKEKI	AEGOWY	APSYV	SPAYHL
Enterobacteriaphage_alpha3_NP_039597/1-235	- GVMMTLAL IRFPPI	S-PLEHHYLA	GKSQLTY	DLAGOP	ALIGNLPF	PREISYRD	LFRD	RS	<mark>GIKI</mark> KV	AESIWY	THPDYY	NFKYHD
EnterobacteriaphageSt-1_YP_002985212/1-235	- GVMMTLMLVEFPPI	S-PLEHHYLV	GRNNLTY	DLAGOP/	ALIGNLPF	REISYOD	LFRD	RP	DTYOK EC	ABSIWY	THPDYY	NYKYOL
CON1900785_F-W1-235	GYLFGIMSTIPEPY	- YOGIDIEW			PSFARLGE	DELUKSE	LCVGISU	140	NTEAD FO	TAPHYAEVI	ABOT	
contio00630 E-0/1-211	GWLIGILSIVPDAL	FOGUNEMM	NR-LNO-F	NY YW	PAFAHLOE	OPIL KSE		E.	NSEGKLEG	VAPPYAEV	VSPST	HGEFFTS
Spiroplasma phage4/1-241	- SYLIVLAVVRYKHT	- DOGIEADW	FRGODK-F	DMYD	PLLANIS	EQPVE-CE	IMVQ	NS	QNEIFC	FOEAWADL	FKPNSV	AGVMRSS
contig00760_F-C/1-240	- GY I IGLAVVRVEHS	Y - DOGL TTKW	TR - GGR - F	TYYD	PRLAALGE	EQPVYNRE	1 YAD	SE	KDSQIFC	YQEAWADY	YKPSYV	TGEMRSN
contig00491_F-D/1-239	- GYLFGLMVI RYDHS	Y - SOGLAPFW	TP - NTF - 1	DFYN	PKFAHLGE	EVPIYKAE	1FAS	ET	IADKTKVFG	YQEIWADY	YRPNMY	TGEMPPG
contig00850_F-A/1-239	GFIIGVCCVPHDHT	Y - QOGLERME	SP-KNK-L	DYYF	PVFANLGE	EQAVERKE	LYAQ	E	TONEAFG	YQECWADY	MKPNR	CGAFRSN
contig00848_F-A/1-239	- BFVIGVCCVRHBHT	- QQGLEMME	- KNK-L	BYYE	PVFANLGE	EQAVERRE	LYAQ	0	TONEAFC	YQEAWADY	MKPNR	CGAFRSN
contig00061 E-0/1-238	GELLIGYMC LEHNES	- DOGLEREW	SD-SDD-L	BYYE	POFANLO	OPIKKKE	LML T	TN	TONETEC	YOFAWADY	MKPNR	SGKMASN
contia00543 F-E/1-238	GELIGVLOVBHDHS	- OOGLERFW	SR-SOR-L	BVVV	POFANLO	OPVICINE	IML T	1.A	TODETEC	YGEAWADY	MKPNRY	SCKMESN
Bdellovibrio_phage_phiMH2K/1-239	- GYVLGF I RARGOVT	Y - QQGLHKMM	SR-QTR-W	DEFW	PKFQELGE	EQAILNKE	1YAQ	NA	·····TOSEIFC	YQERYGEY	FRPSE	KGQFRSN
ChlamydiaphageChp1/1-243	- GEVIGLLSATADLT	Y - QOGLEROW	SR - F SR - Y	DYYW	PTFAHLGE	EQPVYNKE	INCOSDIVADPS	SA	VNDVPFC	YQERYAEY	MKPSKV	TGLFRSN
ChlamydophilapsiltaciphageChp1/1-243	- GEVIGLUSATADLT	Y - QOGLEROW	SR - FSR - Y	BANK	PTFAHLGE	EQPVYNKE	YCOSDTVMDPS	5A	·····VNDVPFC	YQERYAEY	YKPSK	GLERSN
CORRIGO0905_F-B/1-230	CELLOVAVA DYDET			DEMAK	DVEANIGE	EQAILINE	1 VIO	TA	EDDEVEC	YOEAWADY	NUDNDA	TREMOCO
contio00287 F-E/1-239	GEVIGVMVARYDHT	- OOGLEREW	SP-KTR-F	DYWI	PYLANIGE	EQAVLNKE	1YAQ	TA	EDDEVEC	YOEAWGDY	KENRY	TGEMESO
contig00847_F-A/1-239	- GFIIGVMVARYDHT	- QOGLERFW	SR - KDR - F	DFYW	PVFANIGE	CAILNKE	IYAT	KD	·····SDSEVFG	YQEAWADY	YKPSRV	SGEMESN
contig00907_F-B/1-252	- GFLIGVMVARYRHT	Y - QQGLERFW	SR - KDR - F	DYYF	PVFANIGE	EQAINNE	1 YAQ	TV	KDDEVFG	YQEAWADY	YRPNRY	TGEMRSS
contig00288_F-C/1-239	- GELIGVMVARYRHT	Y - QOGLERFW	SR - KDR - F	DYYF	PVFANIGE	EQAIKNEE	1YAQ	TV	KDDEVFG	YQEAWADY	YRPNRY	TGEMESS
contig00401_F-D/1-239	- GYVLGLACVETEHT		N-F-C-	DFYW	PALANIGE	EQAILNKE	1 YLQ	VK	SMNEEAFO	YQEAWAEY	MKPSRV	SSAFESN
CONTROSTU_ANTATCHC_NAKE/T-209	- GUL TGLMSVEADL T			DE MA	PALSHLGE	CAVENEE	NYU	A	TROOVED	VOEBVAEV	PPSL	TOUMOGN
contin00759 F=C/1-235	GVVEGEVNUBARI T	ODGL NPMM		DE W	PVLAHLGE	FOAVL NKE	VA0	-	DDDKVFG	VOERVAEVE	VYPGO	TOKEDOT
contig00213 F-E/1-235	- GYVEGEVNVRADLT	- OOGL NEMW	SR-OGR-F	DFYW	PVLAHLGE	EQAVLNKE	I YAQ	T	DDDKVFG	YOERYAEY	YYPGOL	TGKFRST
ChlamydlaphageChp2/1-243	- GVILGLASVRADLN	Y - DOGLDRMW	SR-RTR-W	DFYW	PALSHLGE	EQAVLNKE	IYCQ(PSVK	NSGGEIVDDOVFG	YOERFAEYF	YKTSK	TGKFRSN
Chlamydiaphage3/1-243	- GVILGLASVRADLN	Y - QQGLDRMM	SR-RTR-W	OFYW	PALSHLGE	EQAVLNKE	1 YCQ (PSVKNS	GGE I VDEQVEG	YQERFAEY	MKTSK	TGKFRSN
Chlamydiaphage4/1-245	- GVILGLASVRADLN	Y - QOGLDRMW	SR-RTR-W	OFYW	PALSHLGE	EQAVENKE	1 YCQ	PAVKDA	QNGNVVVDEQVFG	YQERFAEY	YKTSK	TGKFRSN
ChiamydiaphageCPAROW1-245	GUILGLASVHADLN	- DOGLOHMW			PALSHLGE	EDEVINE	NEO.	AVKDA		YOERFAET	KESK	TOWNER
D4 Highborne Cav/1-227	- CVVIGIVSARADLT	Y-QOGLNRMW	SR-STR-W	DEYW	PALAHIGE	EQAVLNKE	1 YAQ	S	ADDOVEG	YQERFAEY	YKPSL	TGLMRSN
A3_Highborne_Cay/1-234	- CVVIGIVSARADLT	Y - QOGL NRMW	SR-STR-W	DFYW	PALAHIGE	EQAVENKE	IYAQ	NT	DDDEVFG	YOERFAEY	YKPSLT	TGLMRSN
B4_Highborne_Cay/1-234	- CVVIGIVSARADLT	Y - QQGL NRMW	SR-STR-W	VDFYWK	PALAHIGE	EQAVL NKE	IYAQ(N T	D <mark>DD</mark> E <mark>VFG</mark>	YOERFAEY	YKPSL	TGLMRSN
A1_Highborne_Cay/1-234	- CVVIGIVSARADLT	Y - DOGL NRMW	SR-STR-W	PFYW	PALAHIGE	EQAVENKE	I YAQ	T S	ADDDVFG	YOERFAEY	YKPSL	TGLMRSN
83_Highborne_Cay/1-234 Rf_Highborne_Cay/1-234	- CVVIGIVSARADLI				PALAHIGE	EQAVENKE	U CO	S	ADDOVED	VOEDEAEVE	WHEEL	
A2 Highborne Cav/1-234	CVVIGIVSARADLT	Y - OOGL NRMW	SP-STR-V	DEYW	PALAHIGE	FOAVLNKE	YAO-	а	ADDEVEC	YOFREAFY	YKPSLT	TGLMBSN
A4_Highborne_Cay/1-234	- CVVIGIVSARADLT	Y - QQGL NRMW	SR-STR-W	DFYW	PALAHIGE	EQAVENKE	1YAQ(TS	ADDOVEC	YOERFAEY	YKPSLT	TGLMRSN
Contig516_Antarctic_lake/1-241	- GVIIGLVNVRADLT	Y - QQGL S <mark>R</mark> MW	SR - STR - Y	DFYF	PAFATLGE	EQTVLNKE	1 YV T (TS	T <mark>D</mark> NN <mark>VF</mark> G	YQERWAEY	YYPSR	SSLFRST
Contig515_Antarctic_lake/1-241	- GVIIGLVNVRADVT	Y - QQGL SRMW	NR-STR-Y	DFYF	PAFATLGE	EQAVENKE	I YV T (TS	TDNDVFG	YQERWAEY	YYPSR I	SSLFRST
GOS_7913/1-241	- GYVIGLISVRADLT		SR-STR-Y	DEVE	PAFAMLGE	EQAVENKE	FVD	SA	ADSNVFG	YDERWAEY	NPSEV	IGLERST
GOS_111621-241	GVVIGLAAIDADLT			DEVE	PVFAHL GE	EDAVLNE	I VVT	3	TODOVE	VOEDWOEV	VHPAON	TSIFEST
IGOS 6000/1-236	GVVIGLAAIRADLT	Y-QOGLREMM	SR-RTR-Y	DEYE	PVFAHLGE	EQAVENKE	LYVT	TS	TDDDVFG	YQERWAEY	YHPAQY	TSLFRST
GOS_10986/1-241	- GIVLGLVSIRADLT	Y - QQGMRRMW	SR-RTR-Y	DFYF	PVFAHLGE	EQAVENKE	IYT	TS	SDDGVFC	YQERWAEY	YLPAQ	TSLFRST
/GOS_10596/1-241	- GIVLGLVA I RADL T	Y - QQGMR <mark>R</mark> MW	S <mark>R</mark> - RT <mark>R</mark> - Y	DFYF	PVFAHLGE	EQAVENKE	IYT	TS	S <mark>DD</mark> DVFG	YQERWAEY	YLPAQ	TSLFRST
GOS_10933/1-241	- GIVLGLVSIRADLT	Y - QQGMRRMW	SR-RTR-Y	DFYF	PVFAHLGE	EQAVENKE	I Y I T	15	SDDDVFG	YQERWAEY	YLPAQ	TSLFRST
GOS_10391/1-241	- GIVLGLVSIMADLT			BEVE	PVFAHLGE	EQAVENKE	1	5	SDDDVFC	YGERWAEY	YLPAQI	TELERST
GOS 10618/1-241	GIVIGIVSIDADIT			DEVE	PVFAHL GE	EQAVENKE	INT.	S	SDDDVFC	YOERWAEY	VI PAOL	TSI FRST
GOS_7519/1-234	GILLGLVSIRADLT	Y - QOGMRRMM	SR-RTR-Y	DFYF	PVFAHLGE	EQAVLNKE	IYT	TS	SDDDVFC	YQERWAEY	YLPAQI	TSLFRST
Prevotella_bergensis_DSM_17361_(ZP_06006771)/1-248	- GVLMCIYSLVPQIQ	VDCTRLDPMV	DK-LDR-F	DYFTI	PEFENLON	MOPLNSSY	155F(T	DPKNPVLC	YOPRYSEY	TALDVN	IHGOFAQS
Prevotella_buccalis_ATCC_35310_(ZP_06286058)/1-250	- GILMCIYSLVPDMQ	DATRIDPFV	TK-LSR-G	DFFM	PEFEDLGN		1 SD F	2 1 Q	TEKFKC	MOPRYSEY	TSLDIN	IHGQFANG
Prevotella_sp_oral_taxon_317_str_F0108_(ZP_06421583)/1-25	50 - GVLMC I YSVVPAMO	DCMRLDPFV	A <mark>M</mark> - Q IN - G	DYFI	PEFENLON	MOPIVPAF	VSLN		RAKDNSYG	MOPRYSEY	AFDIN	HOOFANG
Dacteroides_plebelus_DSM_17135/1-259 Rectaroides_expectfbil_DSM_20507/4-264		TTOM DRAF	L - TNV - L	DFPI	PEEDRICA			rk <mark>P</mark> Kl	EANASEL VIC	UC BUT NW	TSYDO	MGEFRRS
Bacteroides/1-254	YGLIMCIYHSLPIID	TTDLVNPAF	TK-INS-	DEAL	PEEDRYON	MESVPLVS	LMNP	05	SYNVGSSU	YAPRYISY	TOVDS	
contia00744 F-B/1-253	- GILMCIYHAVPVLD	LRSCODLOL	LH-TLA-	DLPV	PEFDHIGN	EALPIES	LENNQSTE	LA	LMNNIPVLO	YVPRYVAY	TSVDW	SGAFETT
contig00335_F-D/1-257	-GIIMGIYHNVPLLD	YVITGOPONLI	LY-TNT-	DLPF	PEYDSIGN	MOTIOFOR	EVNS - KAVGWTS	VD	YRTQTMC	YLPRFFDV	TRYDEV	LGAF RST
cont/g00377_F-D/1-257	-GIIIGIYHNVPLLD	VITGOPONLI	LY-TNT-A	DLPF	PEFDSIGN		FVNS - KSVSWTS	VD	YRVQ <mark>TM</mark> G	YLPRFFDV	TRYDEY	LGAFRST
contig00845_F-A/1-253	- GVIMCIYHATPLLD	VETGPDLOL	MS - TVKGE	SWPV	PEFDSLGN	MESLPMLS	LVNSKAI	01	VARSYAG	YVPRY I SW	SIDVV	RGAFTDT
contiguo4a9_1+L91-255		VISAPURQU	r A - SMN -	SFFV	MELUAIGL				UA <mark>M</mark> I(<mark>IVG</mark>	TLPRYAW	SIUYV	LWARIEKE

		222	0.40	050	000	070	000	200	200	210	000
		230	240	. 20	260	2/0	280	290	300	310	320
EnterobacteriaphagevvA13_YP_512416/1-231		DGF		· · · · · · · · · · · · · · · · · · ·	· • 🐧 • • • • • • • •	····SATPSDD	I KQRVLV		QSMQLAHWNMQ	TKENCT	
EnterobacteriaphageID18_YP_512/96/1-230		DGF			•••	CAL DOTE	NDILERVLV		USMULAHWNMU		VRHMPI
EnterobacteriaphageG4_INF_040078/1-237		DOF				CALPSTE			OSMOLAHWINMU		VYOHNOT T
Enterobacteriaphagenb2_17_3123701-231		FGF				PSGD.		RHHOVDOCE			
Enterobacteriaphage alpha3 NP 039597/1-235		HGF		PELI		PGTSTG	- DNLOEALLY	RHODYDACE	OSOOLLOWNKO.		
EnteropacteriaphageSt-1_YP_002985212/1-235		EGF		PELI	DDA	PGTTSG	DDLOKATLI	DHNDYNACE	OSOCILOWNNO.	ARYNYN	
contig00185 F-4/1-235	<mark>65</mark> 1	ANW		TEA	RKI	SD - PNL -	NAAFLEV		AVQ DEETD-	KELIW	FSND IRAL
contig00475 F-A/1-211		ANW		TF A	RKL	SAA-NL-	NSDFLSV	POINNPE	FA		
contig00639_F-A/1-211		ANW		TE AI	RR <mark>L</mark>	TAANL	NSDFLSV	POINNPE	FA		
Spiroplasma_phage4/1-241	HPQSI	DYW		<mark>HE</mark> A(DH <mark>Y</mark>	· · AQ <mark>LP</mark> KL -	S SE <mark>WL</mark> K <mark>E</mark> I	DYKNVDRTL	LALK ASDNTP		FMF N 👖 I 🐴 E 🖪
contig00760_F-C/1-240	• <mark>Y</mark> QT <mark>S</mark> I	D AW		<mark>HY</mark> A(DD <mark>Y</mark>	· DK <mark>lptl</mark> -	S AEWI QE	GRENIDRI	IAVT SAVSH -	OFLCD	FWFNETWFR
contig00491_F-D/1-239	VQNSI	AYW		NLA	DH <mark>Y</mark>	· · · · TSE <mark>PTL</mark> ·	S DEWIRE	DVSNVDRAL	AVT SDVSN -	QFWAD	I MIRNKCTR
contig00850_F-A/1-239	ATGTI	DSW		· · · · · · · · HYG	DNY	· · · · SNTPAL -	S QEWMKE	GSAEIARTI	AVE NKVNEP		ILVKNTTV
Contig00848_F-A/1-239		DSW		···············	DNY	THE TREPAL	SQEWMKE	SALIARTI	LAVE NTVNEP	GELAD	FLVKNTTV
COMIG00544_F-E/1-239					JNM	NOVOT	SQEWMKE	SAETART	AVE NKVNEP	OFFAD	FLVKNIIV
00///g00001_F-0/1-238		DEW				DIVEL	S QEVMINE		VVQ DEP		
Rdeliavibrio phage ph/MH2K/1-239	FAFS	DVW		HLA	FYF	TVKPSI	- N - KTELES	NTP- LEBSI	VVIRPDYP	DLIGD	EWEDNT HVR
ChlamvdiaphageChp1/1-243	ATGI	DSW		HL S	2NF	ANLPTL-	N ETEIQS	NTP - IDRAL	AVP - DOP	DEICD	FYENYRCIE
ChlamydophilapsittaciphageChp1/1-243	ATGTI	DSW		HL S(QNF	ANLPTL-	N ETFIQS	NTP - IDRAL	AVP DQP	DFICD	FYENYRCIE
contig00905_F-B/1-236	AQQSI	DAW		<mark>HY</mark> A(QD <mark>Y</mark>	· · · · NELPTL -	S TAWMEQ	SNAEMKRIL	ARS DQP	DELAD	FYEMNKTTR
contig00846_F-A/1-241	YAKSI	D IW		HLAI	DD <mark>Y</mark>	SK <mark>lpsl</mark> -	S AEWIQE	DSSTVNRVL	AASDNLAA-	QFFAD	I YVKNLCTR
contig00287_F-E/1-239	Y AQ <mark>S</mark> I	D VW		······	DD <mark>Y</mark>	TK <mark>lpsl</mark> -	S SE <mark>WI</mark> V <mark>E</mark> I	DKINVDRVL	AVT STNAN -	QLFAD	L <mark>Y I</mark> NNQ T T R
contig00847_F-A/1-239	AKTPL	DSW		HLA	DD <mark>Y</mark>	····DVLPTL-	SDSWIRE	ESNNVNRVL	LAVT SEVSN -	QLFCD	L MYQNR T TR
contig00907_F-B/1-252	APQSI	. D VW		HLG	DDY	ESL <mark>PSL</mark> -	SDSWIRE		LAVSDNVSAQLF	RHLCK	KSLHSTDALVFDSWPYRS
contig00288_F-C/1-239	APQSL			HLG	DDY	· · · · ESLPSL -	SDSWIRE		LAVSDNVSA-	QLFCD	IMVRNLCTR
Contig00401_F-D/1-239	VIGSL	DAW		HYAL	DYY	EAMPVL-	SAEWIQE	TYKNVDRT	AVQSTLED-		FWFKCKCT
COR 405004 225		D AN				SEVOVI					UNEDW/CT
optio00750 E-C/1-235	DROP	D CW					S. POFIOD	NPP VEDVI	LAVO NEP	OFL FD	
contig00213_E_E(1-235	DPOP	nsw.		HIA	DNE	SSIPTI	S. SOFION	NPP VERVI	AVO. NEP	OFLED	STIPLKCAR
ChlamydianhaoeCho?/1-243	ATSS	D SW		HI A(ENIPTI-	S PEELEE		AVS TEP	DELLD	GWE SL RCAR
Chlamydiaphage3/1-243	ATS	DSW		HLA	DEF	ENLPTL-	SPEFIEE		AVSNEP	HFLLD	GWF SLRCAR
Chlamydiaphage4/1-245	ATSSI	DSW		<mark>HL</mark> A(DEF	ENLPTL-	S PEFIEE		AVNTEP	DFLLD	GWFSLRCAR
ChlamydiaphageCPAR39/1-245	ATGSI	DAW		<mark>HL</mark> A(20 <mark>F</mark>	· ENLPTL -	• • S • • PEFIE	NPP-MDRVV	VAVD TEP	DFLLD	GWF SLRCAR
GOS_8011/1-205	AATPL	DHW		HLG	9EF	· · · · ES <mark>LPTL</mark> · ·	🧟 AT 🗾 🗸 🔲	OPP-IDRTI	IAVP SEP	HFILD	S <mark>MI</mark> DMK TT
D4_Highborne_Cay/1-227	ATTSI	DTW				· · · · SALPAL -	N AAF IQE	DPPRPRHCC	PIR		TSLLVROLL
A3_Highborne_Cay/1-234	ATTSL	DTW			VDF	····STLPAL -	NAAFIQE	DPP-VDRVI	AVPSEP	HELED	SYFQYRCA
154_Highborne_Cay/1-234		0 TW			VDF	CAL PAL -	NAAFTUE		AVP SEP	HELED	STECTROARDA
P) Hishborne Cavit-234	TTO	D TW				SAL DAL			AVP SEP		SVEOVDCAE
R1 Habborne Cav/1-233	TTS	DTW.		HIG		SAL PAL	N AAE LOE		AVP SEP	HELED	SVEOVRCAR
A2 Highborne Cav/1-234	ATTS	DTW		HLG	VDF	STL PAL	N AAF LOF		AVP SEP	HELED	SYECYRCAR
A4 Highborne Cay/1-234	ATTSI	DTW		HLG	VDF	· · · · STLPAL ·	N AAFIQE	DPP-VDRVI	IAVP SEP	HELED	- · · · · · · · · · · SYFQYRCAR
Contig516 Antarctic lake/1-241	AAGT	DGW		HLA	QKF	TT <mark>LPTL</mark> -	N T T F I T E	NPP-VSRVL	AVG AAANGQ	QFIFD	SFFDVRKTR
Contig515_Antarctic_lake/1-241	AAGT	DGW		<mark>HL</mark> A(QК <mark>F</mark>	• • • • TT <mark>LPTL</mark> • •	N N T F I T 🧧	NPP-VS <mark>R</mark> VV	VAVG SAANGQ	QFIFD	SFFDVRK <mark>IR</mark>
GOS_7913/1-241	AAGT	DPW		<mark>HY</mark> A(QR <mark>F</mark>	· TSR <mark>PTL</mark> -	N STFIQD	TPP - LARNI	LAVG AAANGQ	QLLLD	AFFD I SAA <mark>R</mark>
GOS_11182/1-241	AAGTI	DAW		HLA(QNF	· · · · SSLPTL -	NTSFIQD	NPP-VDRVV	VAIG AEANGQ	QFIFD	SFFDIKAAR
IGOS_7850/1-241	ASGTL	DAW		HLA	QNF	· · · · SALPTL ·	···S···DTFIVD	QPP - VDRVV	VAIG SGANGQ	QFIFD	SFEDIKAA
GOS_6000/1-236		DAW				ATLET		TOD VORVI	ALC AEANOO	OFIFD	
GOS_109007-247		DAW						TPP-VDPV	ALG. AEANGO	DELED	SEEDIRAA
GOS 10933/1-241		DAW		HI A(ASI PTI	S SSELED		ALG - AFANGO	DELED	SEEDIRAAR
GOS 10391/1-241	AAGTI	DAW		HLA		ATLPTL-	S SSELED	TPP-VDRVV	AIG AEANGO	OFIED	SFED IRAAR
GOS 8425/1-241	AAGTI	DAW		HLA	QNF	····ATLPTL-	SSSFIED	TPP-VDRVV	AIG AEANGQ	QFIFD	SFLIFGQRX
GOS_10618/1-241	AAGTI	DAW		HLA	QNF	····ATLPTL -	S SSFIED	TPP-VDRVV	AIG AEANGQ	QFIFD	SFFDIRAA
GOS_7519/1-234	AAGTL	D AW		HL A(QNF	AT L PTL -	S SSEIED	T PP - VDRVV	VAIGAEANGQ		SX
Prevotella_bergensis_DSM_17361_(ZP_06006771)/1-248	DAL	SSWSV		SRFF	RR <mark>W</mark>	TT <mark>FP</mark> QL	EIADEKII	DPGCLNSIF	PVDYNGTEAND	- CVYGG	CNFNIVKVS
Prevotella_buccalis_ATCC_35310_(ZP_06286058)/1-250	Q <mark>P</mark> I	SYWTV		<mark>G</mark> R <mark>G</mark>	R <mark>AG</mark>	ETLE <mark>TF</mark> - 1	DIA <mark>SL</mark> KII	NPKWLDSIF	FAVNYNGTQITD	- CVFGG	C <mark>QF</mark> N <mark>V</mark> QK <mark>V</mark> S
Prevotella_sp_oral_taxon_317_strF0108_(ZP_06421583)/1-250	0 E <mark>P</mark> l	SYWSI		ARAI	R <mark>GS</mark>	DTLNTF -	N VAALKI	NPHWLDSVF	FAVNYNGTEVTD	CMFGY	AHENTEKVS
Bacteroides_piebelus_DSM_17135/1-259		KTWI I PFD	DEALLA-	ADS'		PDNPNVE	ADSVKAGEEKV	SPSVLDNLF	AVKANSDENTD	FLCS	TLFDVNVV
Dacteroides_eggerthii_DSM_20097/1-261		NSWVISY	NISVLKC	IVTL PNDAPP II		EPVPSVA	MNFTEEKV	NPOCLOPIE	AVQAGODINTD	FLCS	SEEDIKAV
contin00744 E-8/4-253		TW/AD: T	IDOOMT!			CSVECH		TPMVLDSL	VTPCDDTWDTD	DELVH	VAENUUDV
contin00335 F-0/1-257		KNWVAPLO	PANI DC	WLOT			NUT	NPRVIDGI	NVKCDSTIDTO	OFITT	MOLLAV
contig00377_F-D/1-257		KNWVAPLD	- PSYVSk	WLOSSVT	ss <mark>6</mark>	KLALNI	NYGEEKV	NPRVLDS	NYKCDSTIDTD	OFLTA	LYMDIKAVE
contig00845 F-A/1-253		KSWTA		PVD		VIPOEGSPIL-	SYTWEKV	NPSVLNPIP	AVSVDGSWNTD	OLLON	COFDVKVA
contig00489 F-D/1-255	EKEWVAPIT	ELWSNMLK		PLG	тк <mark>с</mark>	TGI	NYNFEKV	NPSILDPIP	AVNADSYWDTD	TFLIN	AFDIRVA
					-						

Fig. S2. Multiple sequence alignment of the capsid protein sequences of the microphages from human faeces, cultured isolates and environmental samples. MUSCLE program (http://www.ebi.ac.uk/Tools/muscle) was used for sequence alignment and aligned partial capsid protein sequences of human faeces, Sagarsso Sea, Highborne Cay, Antarctic lake, *Bacteroides* and *Prevotella* and cultured isolations (Chlamydia-, Bdellovibrio- and Spiroplasma-, Enterobacteria phages) were visualized using Jalview (http://www.jalview.org/).



Fig. S3. Phylogenetic trees of the capsid genes of the microphages from human faeces, environmental samples, *Bacteroides* and *Prevotella* and cultured isolations at the protein (A) and nucleic (B) levels. The sequences were shown in colour as follows: human faeces (red), Sagarsso Sea (deep blue), Highborne Cay (purple), Antarctic lake (light blue), *Bacteroides* and *Prevotella* (gray) and cultured isolations (Chlamydia-, Bdellovibrio- and Spiroplasma phages, green; Enterobacteria phages, black). The phylogenetic trees were constructed based on neighbor-joining algorithm using MEGA 4. Enterobacteria phages were used as outgroups. The scale bars represents 0.2 amino-acid and 0.05 nucleotide substitutions per site, respectively.