Epidemiol. Infect. (2016), **144**, 3549–3553. © Cambridge University Press 2016
doi:10.1017/S095026881600176X

crAssphage is not associated with diarrhoea and has high genetic diversity

Y. Y. LIANG[1], W. ZHANG[1], Y. G. TONG[2] AND S. P. CHEN[1]*

[1] *Department of Laboratory Medicine, Affiliated Hospital of Academy of Military Medical Sciences, Beijing, PR China*
[2] *State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing, PR China*

Received 25 July 2015; Final revision 6 June 2016; Accepted 13 July 2016; first published online 8 August 2016

SUMMARY

crAssphage is a newly discovered gut bacteriophage. However, its pathogenicity and molecular epidemiology in humans are as yet unclear. In this study, we investigated the association between crAssphage and diarrhoea, as well as the molecular epidemiology of crAssphage in Chinese patients from our hospital. Our results indicated that there were no significant differences in the crAssphage-positive ratio and viral loads in faecal supernatants between adults with diarrhoea and healthy adults. Of infants and children with diarrhoea, 2·8% were found to be crAssphage-positive, including two infants aged <1 month. Markedly, of all confirmed crAssphage-positive strains, 100% had the ORF00039 deletion and 77·8% had low identity of ORF00018 compared to crAssphage (GenBank accession no. NC_024711, designated genotype 1). Thus, crAssphage was not associated with diarrhoea and most strains of crAssphage in Chinese patients (designated genotype 2) were characterized by the ORF00039 deletion and low identity of ORF00018.

Key words: crAssphage, diarrhoea, gut bacteriophage, molecular epidemiology.

INTRODUCTION

Bacteriophages are the most abundant biological group on earth [1] and also abundant in the human gut [2]. They specifically infect their bacterial hosts, thereby impacting upon the bacterial ecosystem in human gut microbiota and as a result play an important role in nutrient cycling and carbon flow in biogeochemical and ecological processes. Recent research has demonstrated that the human gut microbiota is involved in obesity, diabetes, metabolic disorders, and diarrhoea, as well as regulatory networks that define good health [3–6]. Therefore, bacteriophages might play a role in some diseases due to the altered microbiota in the human gastrointestinal tract.

A number of gut bacteriophages and interactions between bacteriophages and gut bacteria have been described previously [7, 8]. Recently, a novel bacteriophage, designated crAssphage, which had never been found before, was discovered in human faeces in 2014 [9]. The bacterial host of crAssphage is unknown and predicted to be *Bacteroides* [9]. *Bacteroides* is one of the most numerically prominent genera in the human gut, and bacteriophages can modulate gut microbiota balance in different ways, e.g. transferring bacterial genes [10, 11] and altering bacterial phenotype [12]. As is well known, infections with viral, bacterial, and parasitic pathogens are the most common cause of

* Author for correspondence: Dr S. P. Chen, Department of Laboratory Medicine, Affiliated Hospital of Academy of Military Medical Sciences, Beijing 100071, PR China.
(Email: shpchen@hotmail.com)

https://doi.org/10.1017/S095026881600176X Published online by Cambridge University Press

diarrhoea. Thus, whether crAssphage is associated with diarrhoea is currently in need of investigation.

This study aims to investigate the association between crAssphage and diarrhoea, as well as the molecular epidemiology of crAssphage in Chinese patients.

METHODS

Specimens

This study comprised of 460 cases attending our hospital during the period August 2014 to April 2015, including 327 patients with diarrhoea (249 infants/children, 78 adults) and 133 healthy adults. Faecal samples were collected individually and suspensions of 10% faeces in phosphate buffered saline were prepared immediately. Total DNA was directly extracted from faecal samples, and the supernatant DNA was extracted from faecal supernatants using the QIAamp DNA Mini-kit (Qiagen, Germany).

Polymerase chain reaction (PCR) and sequencing

Two main enzyme-encoding genes (ORF00018: DNA polymerase; ORF00039: endonuclease) and an ORF covering the CRISPR spacer (ORF00053) were chosen as amplifying targets in this study. All primers used are listed in Table 1. Primers For53 and Rev53 were used as detection primers to identify crAssphage infections and primers to amplify partial length of ORF00053 covering the protospacer sequence. Primers For18-P/Rev18-P and For18-F/Rev18-F were used to amplify the partial-length and full-length ORF00018 gene, respectively. Primers For39-F (located within ORF00037) and Rev39-F (located within ORF00040) were used to amplify the gene cluster covering ORF00037, ORF00038, ORF00039, and ORF00040 to verify the deletion of ORF00039. Primers For39-P and Rev39-P, located within ORF00039, were used to confirm whether ORF00039 existed in other positions of the viral genome. The cycling conditions were 95 °C for 1 min, followed by 35 cycles at 95 °C for 10 s, 50 °C for 30 s and 72 °C for 100 s. All PCR products were purified with the QIAquick PCR purification kit (Qiagen) and sequenced using an Applied Biosystems 3700 Genetic Analyzer (Life Technologies, USA). RT-For and RT-Rev were used in a semi-quantitative qPCR assay (SYBR Green dye) to detect crAssphage in human faeces. The qPCR conditions were 95 °C for 1 min, followed by 45 cycles at 95 °C for 10 s, 55 °C for 30 s and 72 °C for 30 s.

Sequence analysis

The sequence cohorts were aligned with ClustalW. The nucleotide and amino-acid homology analysis were performed by using MEGA6 software package [13]. The phylogenetic tree, based on the amplified sequence, was constructed by the maximum-likelihood (ML) method with the Kimura two-parameter settings incorporated in MEGA6.

Statistical analysis

All statistical analysis was performed using the SPSS software package v. 20.0 (IBM Corp., USA) The proportion between adults with diarrhoea and healthy adults was compared using Pearson's χ^2 test. Cycle threshold (Ct) values between the same groups were compared using independent t test. All statistical tests were two-sided, and $P < 0{\cdot}05$ was considered statistically significant.

RESULTS

Prevalence of crAssphage in Chinese patients

As for the adults with diarrhoea and healthy adults, there was no significant difference in the crAssphage-positive ratio [11·5% (9/78) *vs.* 8·3% (11/133), $P > 0{\cdot}05$] and viral loads in faecal supernatants between the two groups (Ct value: $29{\cdot}7 \pm 0{\cdot}5$ *vs.* $29{\cdot}9 \pm 0{\cdot}6$, $P > 0{\cdot}05$), which indicated that crAssphage was not associated with diarrhoea. However, 3·0% (5/166) of infants with diarrhoea (<1 year), 4·8% (4/83) of children with diarrhoea (>1 year), and 11·5% (9/78) of adults with diarrhoea showed a positive correlation between prevalence and age ($P > 0{\cdot}05$). Markedly, crAssphage was detected in two infants aged <1 month (i.e. aged 12 and 24 days).

crAssphage in faecal samples and supernatants

crAssphage both in faecal samples and supernatants was detected in 70/133 healthy adults. Results indicated that 15/70 adults (21·4%) were crAssphage-positive in faecal samples, of which only 5/15 (33·3%) were crAssphage-positive in faecal supernatants.

The molecular epidemiology of crAssphage

Of the 27 confirmed crAssphage-positive individuals, ORF00039 was completely deleted in all strains (using primer For39-F and Rev39-F). Meanwhile, ORF00039

https://doi.org/10.1017/S095026881600176X Published online by Cambridge University Press

Table 1. *Primers used in this study*

Target	Primer name	Sequences (5'→3')	Position	Full/partial length
ORF00018	For18-F	cggcGGGTTAATCAAAATAGAA	8911-8928	Full length
	Rev18-F	gcggAGAACCCCATTTATTAATAAG	11 310-11 330	
ORF00018	For18-P	TGGAAAGGTAAGAAAACTGAAAGA	9916-9939	Partial length
	Rev18-P	AGCAACAATAGGCATAGAATAACC	10 804-10 827	
ORF00039	For39-F	TGCTATTTGGCAAACTGCTGG	23 445-23 465	Full length
	Rev39-F	CTCCAAATCCTTTGTTTCCACGT	25 800-25 822	
ORF00039	For39-P	AAGAATGTATGGGTAAAATGAGAA	24 839-24 862	Partial length
	Rev39-P	ACTGCTAATACTTGGCTGCTGAT	25 277-25 299	
ORF00053	For53	CAGCAGAAGTCCAATCTTTATCAAG	46 182-46 206	Partial length
	Rev53	GATGATGCTGCTGCAATTACTAACG	46 506-46 530	
ORF00086	RT-For	TACGCCTAACCATTGAGGGC	84 130-84 148	
	RT-Rev	ACTGCTCCCGATGGTGTTAC	84 242-84 261	

Four nucleotides (lowercase letters) were added to the 5'-terminus of For18-F and Rev18-F to increase the GC contents of each primer. RT-For and RT-Rev were used in the SYBR Green qPCR assay. The positions for each primer were according to the genome of crAssphage (NC_024711).

was not detected using another pair of primers (For39-P and Rev39-P) which was located within ORF00039. According to the ML tree based on the partial-length ORF00018 sequences (using primers For18-P and Rev18-P), the 27 strains were divided into two genotypes (designated genotypes 1 and 2) (Fig. 1). Interestingly, 77·8% (21/27) of Chinese strains were found to be located within genotype 2, whereas 22·2% (6/27) were located within genotype 1. Meanwhile, genotype 2 had low nucleotide and amino-acid identity (<90%) based on the full-length ORF00018 (using primers For18-F and Rev18-F) compared to crAssphage (genotype 1, GenBank accession no. NC_024711) (Table 2), which was characterized with low nucleotide identity (<70%) in the region between nt 1170 and nt 1800 of ORF00018 (positioned in crAssphage). The characteristics of 27 crAssphage-positive patients and crAssphage sequences are listed in Supplementary Table S1.

DISCUSSION

crAssphage is a newly discovered and commonly found gut bacteriophage in human faeces [9]. However, its pathogenicity and molecular epidemiology in humans are as yet unclear. *Bacteroides*, especially enterotoxigenic *Bacteroides fragilis*, might be a leading cause of diarrhoea [4, 14]. Interestingly, although the host bacteria of crAssphage is unclear, it is predicted to be *Bacteroides* [9]. Thus, crAssphage might be a cause of diarrhoea by modulating the proliferation of *Bacteroides*. However, in our study, no significant differences were observed in the crAssphage-positive ratio and viral loads in faeces between adults with diarrhoea and healthy adults. It is suggested here that crAssphage is not associated with diarrhoea.

It is unknown how crAssphage is transmitted. In this study, we found that 3·6% of infants and children with diarrhoea were crAssphage-positive, especially infants aged <1 month (12 and 24 days), although this bacteriophage has not been found in very young infants previously [9]. Because faecal samples were not available from the mothers of these infants, the strong evidence of amplification and sequencing of the crAssphage gene was not available to demonstrate the close correlation between infants and their mothers. Our study might suggest a possible maternal–neonatal transmission, although other transmission routes (bottle feeding, breastfeeding, or some other ways) can not be ruled out.

Among all 15 crAssphage-positive faecal samples from healthy adults, only five were crAssphage-positive in faecal supernatants, which suggested that crAssphage might be a lysogenic bacteriophage. *Bacteroides* was previously predicted to be the host cells of crAssphage [9]. Thus, after confirmation of its host cells, further experiments (including infection, detection of the integrase gene, prophage sequencing, etc.) need to be performed to confirm this speculation.

The crAssphage strains described in this study have different genome characteristics compared to the strains in United States. The first is ORF00039 which is homologous to endonuclease. The function of endonuclease in bacteriophages is to bind to DNA junctions, and then cleave DNA [15]. ORF00039 has previously been shown to exist in crAssphage [9]. However, it was completely deleted in all the 27 crAssphage-positive strains in our study. The deletion of ORF00039 might be due

https://doi.org/10.1017/S095026881600176X Published online by Cambridge University Press

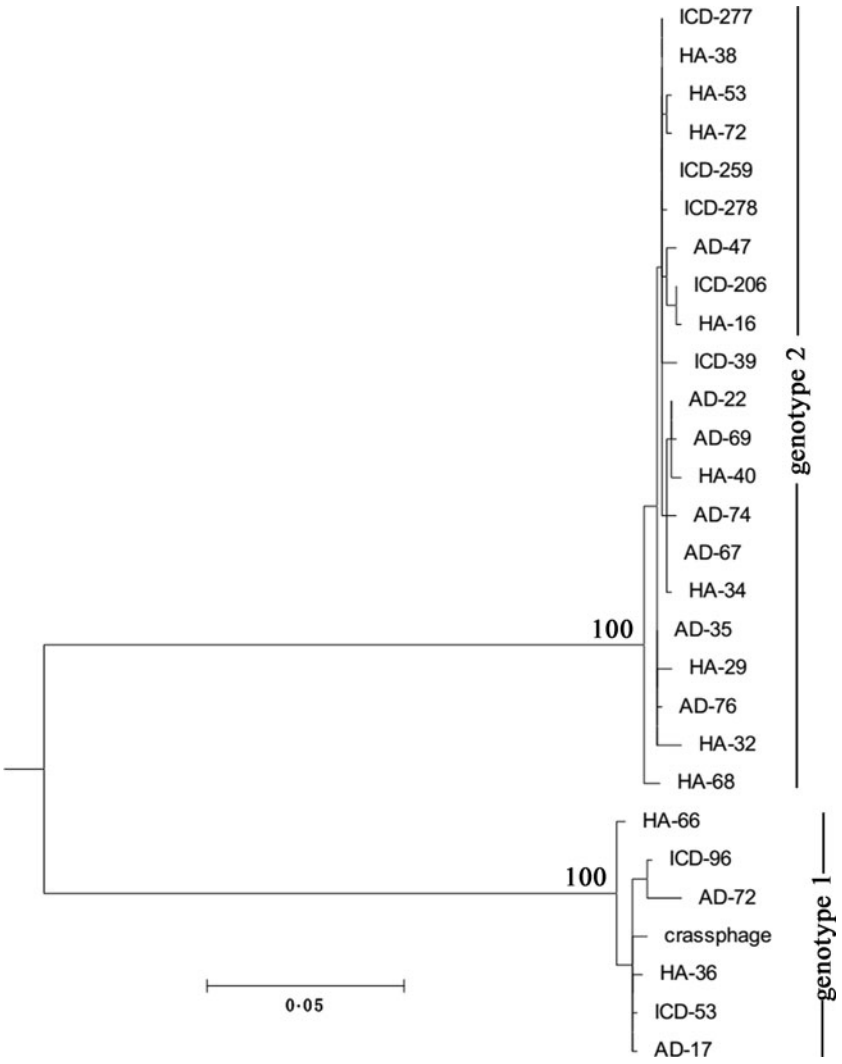

Fig. 1. Phylogenetic analysis by maximum likelihood based on the partial-length ORF00018 gene (846 bp). Bootstrap values in the bootstrap test with 1000 replications were shown on the key nodes of the tree. The scale bar (0·05) indicated substitutions per site. ICD, Infants/children with diarrhoea; AD, adults with diarrhoea; HA, healthy adults. The GenBank accession number of crAssphage is NC_024711.

Table 2. *Nucleotide and amino-acid identity analysis*

	crAssphage	ICD-259	ICD-39	ICD-53	ICD-96	ICD-206	ICD-277	ICD-278
	Amino acid identity (%)							
crAssphage		85·5	85·5	97·5	97·5	85·7	86·0	85·8
ICD-259	86·2		99·1	86·9	87·2	99·3	98·2	98·0
ICD-39	86·2	99·1		86·7	87·0	98·4	98·0	97·9
ICD-53	97·5	87·5	87·4		99·2	87·0	86·0	85·8
ICD-96	97·5	87·8	87·7	99·2		87·0	86·3	86·1
ICD-206	86·3	99·3	98·4	87·7	87·7		98·3	98·2
ICD-277	86·6	98·1	98·0	86·6	86·9	98·3		99·6
ICD-278	86·5	98·0	97·9	86·5	86·8	98·1	99·6	
	Nucleotide identity (%)							

https://doi.org/10.1017/S095026881600176X Published online by Cambridge University Press

to recombination during the replication process of bacteriophages [16] or metaviromic islands [9]. Moreover, its function might be compensated by other proteins, such as ORF00077 (containing a recombinant endonuclease subunit) or counterparts from their bacterial hosts. The second is ORF00018 which encodes polymerase. Our study indicates that the majority (77·8%) of strains described here belong to a very different genotype (genotype 2) characterized by low nucleotide and amino-acid identity (<90%) in ORF00018 compared to crAssphage (genotype 1). The diversities of ORF00039 and ORF00018 presented here are in accord with the previous study [9]. Genes within these variable regions might be under positive evolutionary selection and be involved with host recognition and propagation. The fact that two genotypes (I and II) were simultaneously isolated from different individuals in the same location suggest that these genotypes are stable and durable, and might equalize their host's bacteria populations by a 'kill-the-winner' dynamic.

In conclusion, crAssphage is not associated with diarrhoea and the new genotype is characterized by the ORF00039 deletion and ORF00018 with low identity. However, their host and lysis characterizations as well as their roles in pathogenicity remain unclear and should be investigated in the future.

SUPPLEMENTARY MATERIAL

For supplementary material accompanying this paper visit http://dx.doi.org/10.1017/S095026881600176X.

DECLARATION OF INTEREST

None.

REFERENCES

1. **Hambly E, *et al.*** The viriosphere, diversity, and genetic exchange within phage communities. *Current Opinion in Microbiology* 2005; **8**: 444–450.
2. **Lepage P, *et al.*** A metagenomic insight into our gut's microbiome. *Gut* 2013; **62**: 146–158.
3. **Delzenne NM, *et al.*** Targeting gut microbiota in obesity: effects of prebiotics and probiotics. *Nature Reviews Endocrinology* 2011; **7**: 639–646.
4. **Sack BS, *et al.*** Isolation of entertoxigenic *Bacteroides fragilis* from Bangladeshi children with diarrhea: a controlled study. *Journal of Clinical Microbiology* 1994; **32**: 960–963.
5. **Kimura I, *et al.*** The gut microbiota suppresses insulin-mediated fat accumulation via the short-chain fatty acid receptor GPR43. *Nature Communications* 2013; **4**: 1829.
6. **Hansen TH, *et al.*** The gut microbiome in cardiometabolic health. *Genome Medicine* 2015; **7**: 33.
7. **Waller AS, *et al.*** Classification and quantification of bateriophage taxa in human gut metagenomes. *ISME Journal* 2014; **8**: 1391–1402.
8. **Wagner J, *et al.*** Bacteriophages in gut samples from pediatric Crohn's disease patients: metagenomic analysis using 454 pyrosequencing. *Inflammatory Bowel Diseases* 2013; **19**: 1598–1608.
9. **Dutilh BE, *et al.*** A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nature Communications* 2014; **5**: 4498.
10. **Penadés JR, *et al.*** Bacteriophage-mediated spread of bacterial virulence genes. *Current Opinion in Microbiology* 2015; **23**: 171–178.
11. **Davis BM, *et al.*** Filamentous phages linked to virulence of *Vibrio cholerae*. *Current Opinion in Microbiology* 2003; **6**: 35–42.
12. **Carrolo M, *et al.*** Prophage spontaneous activation promotes DNA release enhancing biofilm formation in *Streptococcus pneumoniae*. *PLoS ONE* 2010; **5**: e15678.
13. **Tamura K, *et al.*** MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology Evolution* 2013; **30**: 2725–2729.
14. **Wick EC, *et al.*** *Bacteroides* spp. and diarrhea. *Current Opinion in Infectious Diseases* 2010; **23**: 470–474.
15. **Freeman AD, *et al.*** The importance of the N-terminus of T7 endonuclease I in the interaction with DNA junctions. *Journal of Molecular Biology* 2013; **425**: 395–410.
16. **Nafissi N, *et al.*** Bacteriophage recombination systems and biotechnical application. *Applied Microbiology and Biotechnology* 2014; **98**: 2841–2851.

https://doi.org/10.1017/S095026881600176X Published online by Cambridge University Press