

DATA NOTE

Open Access



# Whole-genome of Mexican-crAssphage isolated from the human gut microbiome

Melany Cervantes-Echeverría<sup>1</sup>, Edgar Equihua-Medina<sup>1</sup>, Fernanda Cornejo-Granados<sup>1</sup>, Abigail Hernández-Reyna<sup>1</sup>, Filiberto Sánchez<sup>1</sup>, Blanca Estela López-Contreras<sup>2</sup>, Samuel Canizales-Quinteros<sup>2</sup> and Adrián Ochoa-Leyva<sup>1\*</sup>

## Abstract

**Objectives:** crAssphage is a newly found phage described as the most abundant virus in the human gut microbiome. The majority of the crAssphage proteins are unknown in sequences databases, and its pathogenicity and epidemiology in humans are yet unclear. Hence, being one of the most abundant phages in the human gut microbiome more investigation at the genomic level is necessary to improve our understanding, especially in the Latin American population.

**Data description:** In this article, we provide the whole genome of a crAssphage isolated from the human gut microbiome of the Mexican population, which was named Mexican-crAssphage. The genome consists of 96,283 bp, G+C content of 29.24% and 87 coding sequences. Notably, we did not find any transfer RNA genes in the genome sequence. We also sequenced viral-like enriched particles from 28 fecal samples, and we detected the presence of the Mexican-crAssphage genome in 8 samples (28.5%). To our knowledge, our data is the first whole genome report of the crAssphage isolated from the Latin American Population and provides valuable information for the experimental characterization of the most abundant human gut bacteriophage. The whole genome shotgun project of the Mexican-crAssphage is available at DDBJ/ENA/GenBank under the GenBank MK069403.

**Keywords:** crAssphage, Mexican-crAssphage, Human gut microbiome, Human phages, Host-microbiota

## Objective

The human body is inhabited by a high diversity of bacteria, archaea, fungi, protozoa, and viruses. These microbes are collectively known as the human microbiota, whereas their collective genomes form the human microbiome [1]. The human gut virome is dominated by bacteriophages [2], infecting their bacterial hosts and they also impact the microbiome composition [1]. Interestingly, it has been proposed that bacteriophages may have a role in shaping the diversity and composition of the microbiota [1] and also play a role in some diseases such as bowel disease [3] and type 1 and 2 diabetes [4, 5].

A novel bacteriophage, named crAssphage, was recently discovered as the most abundant virus in the human gut microbiome [6]. After that, a crAss-like family was discovered and appears to be abundant and widespread in diverse habitats, both animal and environmental associated [7]. Various bacteria of the phylum Bacteroidetes appear to be the primary hosts of crAss-like phages [6, 7]. For example,  $\Phi$ CrAss001, isolated from the human feces, was the first member of the extensive crAssphage family to be grown in pure culture and this phage infects the human gut symbiont *Bacteroides intestinalis* [8]. Recently, 98 complete circular genomes of crAss-like phages were reported and helped to establish the classification of this phage family into four candidate subfamilies composed of 10 candidate genera [9]. Furthermore, crAssphage was not associated with diarrhea in Chinese patients [10]. The crAssphage genomes have been isolated from the human gut of several geographical

\*Correspondence: aochoa@ibt.unam.mx

<sup>1</sup> Departamento de Microbiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Avenida Universidad 2001, Colonia Chamilpa, 62210 Cuernavaca, Morelos, Mexico  
Full list of author information is available at the end of the article



origins (Data file 1 in Table 1). However, a genome sequence from this phage family has not been reported to date in a Latin American population. Hence, being one of the most abundant phages in the human gut microbiome more investigation at the genomic level is necessary to improve our understanding about their function, especially in the Latin American population.

### Data description

Phage-enriched filtrates of fecal samples from 28 Mexican children were isolated using a modified protocol [11]. In brief, 250 mg of feces were homogenized in SM Buffer for each sample and centrifuged 30 min at 4700×g. The supernatant was filtered through a 0.22 µm PES filter (720–1320, Nalgene, USA) and concentrated in Amicon Ultra 15, 100KDa (UFC910096, Millipore, USA). Then, Amicon was washed using one volume of SM Buffer, and the viral particles were concentrated in 200 µl of SM buffer. We extracted the DNA using the QIAamp MinElute Virus Spin kit (57704, QIAGEN, Hilden, Germany). The DNA quality and quantity were measured using agarose gel electrophoresis and Qubit High-sensitivity fluorometric assay (Cat. Q32851, Life Technologies, Carlsbad, CA, USA), respectively. The DNA was used to construct the pair-end libraries using the Nextera XT DNA Library Preparation kit (Cat. FC-131-1024, Illumina, CA, USA) selecting an insert size of 400–600 bp with the Ampure XP beads (Cat. A63882, Beckman Coulter, CA, USA). The libraries were analyzed with the 2100 Bioanalyzer instrument (Cat. 5067-1504, Agilent Technologies, CA, USA), and sequencing was performed using the Illumina NextSeq500 with a 300 cycle paired-end format (FC-404-2003; Illumina, CA, USA) at the National Institute of Genomic Medicine in Mexico City. The reads were analyzed using FastQC version 0.11.5, and only the reads with a quality > Q20 were used for further analysis. The resulting reads from each sample were mapped against the crAssphage reference genome (GenBank ID: JQ995537) using SMALT version 0.7.6. After that, we selected the sample with the highest number of reads mapped to crAssphage genome to conduct a

denovo genome assembly using Spades version 3.8.1. The resulting contigs were ordered using MeDuSa [12] setting the default parameters.

The total size of the assembled Mexican-crAssphage genome was 96,283 bp and G+C content of 29.24% (Data set 1 in Table 1). The reads coverage of our Mexican-crAssphage genome was 188X. To visualize the read depth and codified proteins along the Mexican-crAssphage genome we used DNAPlotter (Data file 2 in Table 1). A total of 87 coding sequences (CDS) were predicted using RAST [13]. They were largely co-oriented, organized in two blocks of CDS alongside the genome. These sequences were BLASTed against the non-redundant (NR) proteins database using Blast2GO [14]. After that, 12 proteins (13.8%) were unknown, and 60 proteins (68.9%) were defined as hypothetical protein. The genome showed to encode phage proteins, including proteins involved in nucleic acid manipulation (helicase, ligase, primase, and polymerase), and phage structural proteins. Notably, we did not find any transfer RNA genes in the genome sequence.

Finally, the viral reads of the 28 samples were mapped against the Mexican-crAssphage genome using SMALT version 0.7.6. We detected the presence of this phage's genome in eight samples, meaning that the Mexican-crAssphage was present in 28.5% of the analyzed samples. This is the first CrAssphage genome isolated from a Latin-American population, and it can be used in different applications of human viral metagenomics to understand the impact that host-genetics have in modulating the evolution of crAssphage across the world.

### Limitations

A more significant deep sequencing of viral particles should be used in the future to improve a region of 100 uncalled bases (N's) reported in this Mexican-crAssphage genome. This region is in positions 40,116–40,215 of the reported genome. It is important to note that these are the only missing bases from all the genome.

**Table 1 Overview of data files/data sets**

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data set 1	Whole genome	FASTA file	GenBank Accession number: MK069403 ( <a href="https://www.ncbi.nlm.nih.gov/nucleotide/MK069403">https://www.ncbi.nlm.nih.gov/nucleotide/MK069403</a> )
Data file 1	Geographical origin of the crAssphage genomes previously reported	MS Excel file (.xlsx)	Figshare ( <a href="https://doi.org/10.6084/m9.figshare.7379600">https://doi.org/10.6084/m9.figshare.7379600</a> )
Data file 2	Read depth and proteins along the Mexican-crAssphage genome.	Adobe Portable Document Format (.pdf)	Figshare ( <a href="https://doi.org/10.6084/m9.figshare.7379603">https://doi.org/10.6084/m9.figshare.7379603</a> )

**Authors' contributions**

MCE, EEM, and AOL conceived and designed the experiments, performed the experiments, analyzed the data. FCG, AHR, FS, and AOL performed the experiments, analyzed the data, and contributed the reagents/materials/analysis tools. SCQ, BELC and AOL contributed the reagents/materials. All authors read and approved the final manuscript.

**Author details**

<sup>1</sup> Departamento de Microbiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Avenida Universidad 2001, Colonia Chamilpa, 62210 Cuernavaca, Morelos, Mexico. <sup>2</sup> Unidad de Genómica de Poblaciones Aplicada a la Salud, Facultad de Química, UNAM/Instituto Nacional de Medicina Genómica (INMEGEN), Ciudad de México, México.

**Acknowledgements**

We thank Juan Manuel Hurtado-Ramírez and Gamaliel López-Leal of IBT-UNAM for bioinformatics technical support and Dr. Ricardo Alfredo Grande Cano and Gloria Tanahiry Vázquez Castro at IBT-UNAM for their experimental sequencing support. We thank Alfredo Mendoza-Vargas and Unidad de Secuenciación Masiva of the Instituto Nacional de Medicina Genómica for their technical support in the samples sequencing.

**Competing interests**

The authors declare that they have no competing interests.

**Availability of data materials**

The data described in this Data Note (whole genome project) is freely available at DDBJ/EMBL/GenBank under the accession number PRJNA495080 (<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA495080>) and GenBank MK069403 (<https://www.ncbi.nlm.nih.gov/nucore/MK069403>). The data files 1 and 2 are available at <https://doi.org/10.6084/m9.figshare.7379600> and <https://doi.org/10.6084/m9.figshare.7379603>, respectively.

**Consent for publication**

Not applicable.

**Ethics approval and consent to participate**

The Ethic Committee of the National Institute of Genomic Medicine in Mexico City approved the study. The parents or guardians of donors signed the informed consent form for participation, and the donors assented to participate.

**Funding**

The National Council for Science and Technology of Mexico (CONACYT) funded the reported study, Grant No. SALUD-2014-C01-234188 and the DGAPA PAPPIT UNAM (IA203118).

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 20 October 2018 Accepted: 12 December 2018

Published online: 17 December 2018

**References**

- Bikel S, Valdez-Lara A, Cornejo-Granados F, Rico K, Canizales-Quintero S, Soberón X, Del Pozo-Yauner L, Ochoa-Leyva A. Combining metagenomics, metatranscriptomics and viromics to explore novel microbial interactions: towards a systems-level understanding of human microbiome. *Comput Struct Biotechnol J*. 2015;13:390–401. <https://doi.org/10.1016/j.csbj.2015.06.001>.
- Minot S, Sinha R, Chen J, Li H, Keilbaugh S, Wu G, Lewis J, Bushman F. The human gut virome: Inter-individual variation and dynamic response to diet. *Genome Res*. 2011;21(10):1616–25. <https://doi.org/10.1101/gr.122705.111>.
- Norman J, Handley S, Baldrige M, Doit L, Liu C, Keller B, et al. Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell*. 2015;160(3):447–60. <https://doi.org/10.1016/j.cell.2015.01.002>.
- Ma Y, You X, Mai G, Tokuyasu T, Liu C. A human gut phage catalog correlates the gut phageome with type 2 diabetes. *Microbiome*. 2018;6(1):24. <https://doi.org/10.1186/s40168-018-0410-y>.
- Zhao G, Vatanen T, Droit L, Park A, Kostic AD, Poon TW, et al. Intestinal virome changes precede autoimmunity in type I diabetes-susceptible children. *Proc Natl Acad Sci U S A*. 2017;114(30):E6166–75. <https://doi.org/10.1073/pnas.1706359114>.
- Dutilh B, Cassman N, McNair K, Sanchez S, Silva G, Boling L, et al. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat Commun*. 2014;24(5):4498. <https://doi.org/10.1038/ncomms5498>.
- Yutin N, Makarova K, Gussow A, Krupovic A, Segall A, Edwards R, Koonin E. Discovery of an expansive bacteriophage family that includes the most abundant viruses from the human gut. *Nat Microbiol*. 2018;3:38–46. <https://doi.org/10.1038/s41564-017-0053-y>.
- Shkoporov A, Khokhlova E, Fitzgerald C, Stockdale S, Draper L, Ross R, Hill C. ΦCrAss001, a member of the most abundant bacteriophage family in the human gut, infects Bacteroides. *bioRxiv*. 2018. <https://doi.org/10.1101/354837>.
- Guerin E, Shkoporov A, Stockdale S, Clooney A, Ryan F, Draper L, Gonzalez-Tortuero E, Ross P, Hill C. Biology and taxonomy of crAss-like bacteriophages, the most abundant virus in the human gut. *bioRxiv*. 2018. <https://doi.org/10.1101/295642>.
- Liang YY, Zhang E, Tong YG, Chen SP. CrAssphage is not associated with diarrhoea and has high genetic diversity. *Epidemiol Infect*. 2016;1(16):1–5. <https://doi.org/10.1017/S095026881600176X>.
- Reyes A, Wu M, McNulty NP, Rohwer F, Gordon J. Gnotobiotic mouse model of phage–bacterial host dynamics in the human gut. *Proc Natl Acad Sci USA*. 2013;110(50):20236–41. <https://doi.org/10.1073/pnas.1319470110>.
- Bosi E, Donati E, Galardini M, Brunetti S, Sagot MF, Lió P, Crescenzi P, Fani R, Fondi M. MeDuSa: a multi-draft based scaffold. *Bioinformatics*. 2015;31(15):2443–51. <https://doi.org/10.1093/bioinformatics/btv171>.
- Aziz R, Bartels D, Best A, DeJongh M, Disz T, Edwards R, et al. The RAST server: rapid annotations using subsystems technology. *BMC Genomics*. 2008;9(1):75–90. <https://doi.org/10.1186/1471-2164-9-75>.
- Götz S, García-Gómez J, Terol J, Williams T, Nagaraj S, Nueda M, Robles M, Talón M, Dopazo J, Conesa A. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res*. 2008;36(10):3420–35. <https://doi.org/10.1093/nar/gkn176>.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

