AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

| Bacteriophages | Announcement

Virome assembly reveals draft genomes of native Pseudomonas phages isolated from a paediatric bronchoalveolar lavage sample

Patricia Agudelo-Romero,[1,2,3] Jose A. Caparros-Martin,[1,4,5] Abhinav Sharma,[6] Montserrat Saladié,[4] Peter D. Sly,[7] Stephen M. Stick,[8,9] Fergal O'Gara,[1,4,10] COMBAT study group

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT We present lung virome data recovered through shotgun metagenomics in bronchoalveolar lavage fluid from an infant with cystic fibrosis, who tested positive for *Stenotrophomonas maltophilia* infection. Using a bioinformatic pipeline for virus characterization in shotgun metagenomic data, we identified five viral contigs representing Pseudomonas phages classified as Caudoviricetes.

KEYWORDS viruses, cystic fibrosis, bronchoalveolar lavage, metagenomics

Cystic fibrosis (CF) is a genetic condition that disrupts airway physiology, making patients susceptible to lung infections and chronic inflammation (1). While CF research has characterized lung bacterial communities, the associated virome remains understudied. Characterizing the lung virome is essential for understanding microbiome dynamics and its impact on respiratory health in CF. We present five viral contigs (vContigs) from shotgun metagenomic data in bronchoalveolar lavage fluid (BALF) of a CF infant with a *Stenotrophomonas maltophilia* lung infection (2).

This study represents an exploratory outcome of the COMBAT CF clinical trial (Clinicaltrials.gov: NCT01270074) (3). The COMBAT CF study protocol was approved by site-specific hospital Human Research Ethics Committees (HREC) and the HREC at the University of Western Australia reviewed this study and granted ethics exemption (2024/E000843). For nucleic acid extraction, 2 mL of BALF was centrifuged at 20,000 × *g* for 30 min at 4°C. Pellets underwent enzymatic digestion (MetaPolyzyme and proteinase K), followed by bead-beating, and chloroform:isoamyl alcohol extraction. Nucleic acids were then precipitated from the aqueous phase with polyethylene glycol, centrifuged using the same conditions as above, washed with ethanol, and resuspended in sterile water (2, 4). Libraries were built using the Nextera XT kit (Illumina, San Diego, CA, USA), and sequenced using a 150 bp pair-end configuration in a NovaSeq 6000 (Illumina) instrument at Genewiz (China). We retained 154 million reads with a mean Phred-like *Q*-score greater than or equal to 35 (5). We used the Snakemake pipeline EVEREST-meta for virus discovery (https://github.com/agudeloromero/EVEREST_meta)(6), using the database v4 (November 2024) (7), as we described (8, 9). All tools were run with default parameters unless otherwise specified.

Human reads were removed with minimap2 v2.24 (10), followed by deduplication and digital normalization using BBMAP v38.96 (11). We retained 2,559,801 non-human reads, which were *de novo* assembled with SPAdes v3.13.0 (12). Subsequently, viral contigs longer than 5,000 bp were retained using VirSorter v2.2.3 (13). Quality genome assessment was performed using CheckV v0.9.0 (14). EVEREST-meta uses the Lowest Common Ancestor algorithm from MMSeq2 v13.45111 taxonomy tool (15, 16), using

Editor Kenneth M. Stedman, Portland State University, Portland, Oregon, USA

Address correspondence to Stephen M. Stick, stephen.stick@health.wa.gov.au, or Fergal O'Gara, f.ogara@ucc.ie.

Patricia Agudelo-Romero and Jose A. Caparros-Martin contributed equally to this article. Author order was determined alphabetically.

The authors declare no conflict of interest.

See the funding table on p. 4.

Received 30 September 2024
Accepted 5 December 2024
Published 20 December 2024

TABLE 1 Five viral contigs (vContigs) were recovered from shotgun metagenomic data from DNA from BALF of an infant with cystic fibrosis[a]

Contigs	vContig_1	vContig_2	vContig_3	vContig_4	vContig_5
Genome size (bp)	25,966	14,256	12,689	7,952	5,318
Genome coverage (×)	34.1506	13.8512	38.1961	9.2591	12.422
No. of mapped reads	5,816	1,291	3,193	493	440
GC content (%)	59.43	60.96	60.71	60.48	58.99
CheckV completeness (%)	57.36	27.71	24.79	16.93	11.63
CDS*	39	22	30	10	11
Connector*	3	2	0	0	0
DNA, RNA, and nucleotide metabolism*	2	0	6	1	2
Head and packaging*	6	3	0	0	0
Integration and excision*	1	0	1	0	0
Lysis*	2	1	0	0	0
Moron, auxiliary metabolic gene, and host takeover*	0	0	1	0	0
Other*	0	1	1	0	1
Tail*	8	2	0	6	0
Transcription regulation*	0	1	1	0	1
Unknown function*	17	12	20	3	7
tRNAs, CRISPRs, tmRNAs*	0	0	0	0	0
Virulence factors (VFDB)	0	0	0	0	0
AMR genes (CARD)	0	0	0	0	0
Lowest common ancestor by nucleotide (NCBI) database	Class: Caudoviricetes	Class: Caudoviricetes	Class: Caudoviricetes	Class: Caudoviricetes	Class: Caudoviricetes
Closest-related phage NBCI, nt (GenBank accession #)	Pseudomonas phage PS-1 (NC_029066.1)	Pseudomonas phage PS-1 (NC_029066.1)	Pseudomonas virus D3 (NC_002484.2)	Ralstonia phage Dina (NC_055026.1)	Pseudomonas phage JBD44 (NC_030929.1)
Average nucleotide identity by orthology (OrthoANI) to the closest-related phage	75.74	87.80	71.62	61.36	79.21
Lowest common ancestor by amino acid (UniProt; TrEMBL) database	Class: Caudoviricetes	Class: Caudoviricetes	Class: Caudoviricetes	Class: Caudoviricetes	Class: Caudoviricetes
Closest-related phage UniProt, aa (UniProtKB accession #)	Pseudomonas phage PS-1 (A0A0H5ART3)	Pseudomonas phage PS-1 (A0A0H5AWC7)	Pseudomonas phage vB_PeaS_FBPa47 (A0A9E7QP96)	Pseudomonas phage PMBT14 (A0A2S1B6B3)	Pseudomonas phage JBD44 (A0A125RNK2)
Host prediction, genus level	Pseudomonas	Pseudomonas	Pseudomonas	Pseudomonas	Pseudomonas
Confidence score of host prediction	94	94.9	96.9	93.9	96.3
GenBank accession number	PP986815	PP986816	PP986817	PP986818	PP986819

[a]Functional annotations performed with Pharokka (17), are indicated with an asterisk.

NCBI (nucleotide; nt) and UniProt (amino acid; aa) databases for viral taxonomic classification and to obtain the closest related virus. Complementary analyses were performed for functional annotation with Pharokka v1.3.0 (17), virulence and antibiotic resistance gene identification using ABRICATE v1.0.1 (https://github.com/tseemann/abricate) (18) with the VFDB (July 2019) (19) and CARD (July 2019) (20) databases, average nucleotide identity by orthology (OrthoANI) v0.6.0 (21), and host prediction using iPHoP v1.2.0 (22).

Five vContigs were characterized with a length between 25,966 and 5,318 bp and 60.11% average GC content (Table 1). Notably, no vContigs contained virulence or antimicrobial resistance genes, while the predicted genes were primarily related to DNA, RNA, nucleotide metabolism, head, and packaging biological processes, as well as unknown function (Table 1). The five vContigs were classified as Caudoviricetes

using both NCBI and UniProt databases. Using UniProt, the closest related phage for all vContigs were *Pseudomonas* phages. Host prediction supported *Pseudomonas* as the potential host (Table 1). No phages associated with *S. maltophilia* were detected.

ACKNOWLEDGMENTS

The authors acknowledge the patients and their families for their participation in the COMBAT-CF clinical trial, as well as all members of the COMBAT-CF study team. S.M.S. obtained funding from the US Cystic Fibrosis Foundation to support the COMBAT-CF study (STICK10K0, STICK19K0). The additional support was obtained by S.M.S. from the National Health and Medical Research Council of Australia (NHMRC115648) and the Australian NHMRC 2020 Synergy grant (APP1183640). F.O. received funding to support this work from the Health Research Board (HRB-ILP-POR-2019-004), the Irish Thoracic Society (MRCG-2018-16 and MRCG-2014-6), the US Cystic Fibrosis Foundation (OGARA1710), the Glenn Brown Memorial Grant 2017 (The Institute for Respiratory Health, Perth, Australia), the European Commission (EU-634486), Science Foundation Ireland (SSPC-2/PharM5, 13/TIDA/B2625, 14/TIDA/2438, and 15/TIDA/2977), and Enterprise Ireland Commercialisation Fund (CF-2017-0757 P). P.A.-R. received support from a Google Cloud Education Program grant and the Telethon Kids Institute Theme Collaboration Award 2023.

AUTHOR AFFILIATIONS

[1]Wal-Yan Respiratory Research Centre, Telethon Kids Institute, Perth, Western Australia, Australia
[2]Australian Research Council Centre of Excellence in Plant Energy Biology, School of Molecular Sciences, The University of Western Australia, Perth, Western Australia, Australia
[3]European Virus Bioinformatics Center, Friedrich-Schiller-Universitat Jena, Thuringia, Germany
[4]Curtin Health Innovation Research Institute (CHIRI), Curtin University, Perth, Western Australia, Australia
[5]UWA Medical School, The University of Western Australia, Perth, Western Australia, Australia
[6]DSI-NRF Centre of Excellence for Biomedical Tuberculosis Research, SAMRC Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa
[7]Children's Health and Environment Program, Child Health Research Centre, The University of Queensland, Brisbane, Australia
[8]Department of Respiratory and Sleep Medicine, Perth Children's Hospital, Perth, Western Australia, Australia
[9]Centre for Cell Therapy and Regenerative Medicine, School of Medicine and Pharmacology, The University of Western Australia and Harry Perkins Institute of Medical Research, Perth, Western Australia, Australia
[10]BIOMERIT Research Centre, School of Microbiology, University College Cork, Cork, Ireland

PRESENT ADDRESS

Montserrat Saladié, Eurecat, Centre Tecnològic de Catalunya, Centre for Omic Sciences (COS), Joint Unit Universitat Rovira I Virgili-EURECAT, Reus, Catalonia, Spain

AUTHOR ORCIDs

Jose A. Caparros-Martin http://orcid.org/0000-0003-1214-4952
Stephen M. Stick http://orcid.org/0000-0002-5386-8482
Fergal O'Gara http://orcid.org/0000-0002-2659-0673

FUNDING

Funder	Grant(s)	Author(s)
Cystic Fibrosis Foundation (CFF)	STICK10K0	Stephen M. Stick
European Commission (EC)	EU-634486	Fergal O'Gara
Science Foundation Ireland (SFI)	SSPC-2/PharM5	Fergal O'Gara
Science Foundation Ireland (SFI)	13/TIDA/B2625	Fergal O'Gara
Science Foundation Ireland (SFI)	14/TIDA/2438	Fergal O'Gara
Science Foundation Ireland (SFI)	15/TIDA/2977	Fergal O'Gara
Enterprise Ireland Commercialisation Fund	CF-2017-0757-P	Fergal O'Gara
Google for Education	Google Cloud Education Program grant	Patricia Agudelo-Romero
Telethon Kids Institute Theme Collaboration Award 2023	Theme Collaboration Award 2023	Patricia Agudelo-Romero
Cystic Fibrosis Foundation (CFF)	STICK19K0	Stephen M. Stick
DHAC | National Health and Medical Research Council (NHMRC)	NHMRC115648	Stephen M. Stick
DHAC | National Health and Medical Research Council (NHMRC)	APP1183640	Stephen M. Stick
Cystic Fibrosis Foundation (CFF)	OGARA1710	Fergal O'Gara
Health Research Board (HRB)	HRB-ILP-POR-2019-004	Fergal O'Gara
Irish Thoracic Society (ITS)	MRCG-2018-16	Fergal O'Gara
Irish Thoracic Society (ITS)	MRCG-2014-6	Fergal O'Gara
The Institute for Respiratory Health	Glenn Brown Memorial Grant 2017	Fergal O'Gara

AUTHOR CONTRIBUTIONS

Patricia Agudelo-Romero, Conceptualization, Formal analysis, Investigation, Methodology, Software, Writing – original draft, Writing – review and editing | Jose A. Caparros-Martin, Conceptualization, Formal analysis, Investigation, Methodology, Software, Writing – original draft, Writing – review and editing | Abhinav Sharma, Formal analysis, Software, Writing – review and editing | Montserrat Saladié, Investigation, Methodology, Writing – review and editing | Peter D. Sly, Conceptualization, Funding acquisition, Resources, Supervision, Writing – review and editing | Stephen M. Stick, Conceptualization, Funding acquisition, Resources, Supervision, Writing – review and editing | Fergal O'Gara, Conceptualization, Funding acquisition, Resources, Supervision, Writing – review and editing.

DATA AVAILABILITY

This Whole Genome Shotgun project has been deposited in the BioProject PRJNA1126024; GenBank accession numbers PP986815, PP986816, PP986817, PP986818, and PP986819; BioSample SAMN41940487; Sequence read archive accession number SRR29521294.

REFERENCES

1. Elborn JS. 2016. Cystic fibrosis. Lancet 388:2519–2531. https://doi.org/10.1016/S0140-6736(16)00576-6
2. Caparrós-Martín JA, Saladie M, Agudelo-Romero SP, Reen FJ, Ware RS, Sly PD, Stick SM, O'Gara F, COMBAT study group. 2023. Detection of bile acids in bronchoalveolar lavage fluid defines the inflammatory and microbial landscape of the lower airways in infants with cystic fibrosis. Microbiome 11:132. https://doi.org/10.1186/s40168-023-01543-9
3. Stick SM, Foti A, Ware RS, Tiddens H, Clements BS, Armstrong DS, Selvadurai H, Tai A, Cooper PJ, Byrnes CA, Belessis Y, Wainwright C, Jaffe A, Robinson P, Saiman L, Sly PD. 2022. The effect of azithromycin on structural lung disease in infants with cystic fibrosis (COMBAT CF): a phase 3, randomised, double-blind, placebo-controlled clinical trial. Lancet Respir Med 10:776–784. https://doi.org/10.1016/S2213-2600(22)00165-5

4. Saladié M, Caparrós-Martín JA, Agudelo-Romero P, Wark PAB, Stick SM, O'Gara F. 2020. Microbiomic analysis on low abundant respiratory biomass samples; improved recovery of microbial DNA from bronchoalveolar lavage fluid. Front Microbiol 11:572504. https://doi.org/10.3389/fmicb.2020.572504
5. Andrews S, Krueger F, Segonds-Pichon A, Biggins L, Krueger C, Wingett S. 2012. FastQC: a quality control tool for high throughput sequence data. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc
6. Agudelo-Romero P, Conradie T, Kicic A, Caparros-Martin J, Stick S. 2024. EVEREST-Meta: a pipEline for Viral assEmbly and chaRactEriSaTion for METAgenomics v0.1.0. Available from: https://doi.org/10.5281/zenodo.10487446
7. Agudelo-Romero P, Sharma A, Conradie T, Kicic A, Caparros-Martin JA, Stick SM. 2024. Database for EVEREST (pipEline for Viral assEmbly and chaRactEriSaTion) v0.0.4b. Available from: https://doi.org/10.5281/zenodo.8361918
8. Conradie T, Caparros-Martin JA, Egan S, Kicic A, Koks S, Stick SM, Agudelo-Romero P. 2024. Exploring the Complexity of the Human Respiratory Virome through an In Silico Analysis of Shotgun Metagenomic Data Retrieved from Public Repositories. Viruses 16:953. https://doi.org/10.3390/v16060953
9. Agudelo-Romero P, Caparros-Martin JA, Sharma A, Saladié M, Sly PD, Stick SM, O Gara F, COMBAT study group. 2024. A near-complete genome of the uncultured *Staphylococcus aureus* phage COMBAT-CF_PAR1 isolated from the lungs of an infant with cystic fibrosis. Microbiol Resour Announc:e0104724. https://doi.org/10.1128/mra.01047-24
10. Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191
11. 2016. BBMAP: short-read aligner, and other bioinformatics tools. Available from: http://sourceforge.net/projects/bbmap
12. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using SPAdes *de novo* assembler. Curr Protoc Bioinformatics 70:e102. https://doi.org/10.1002/cpbi.102
13. Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, Pratama AA, Gazitúa MC, Vik D, Sullivan MB, Roux S. 2021. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. Microbiome 9:37. https://doi.org/10.1186/s40168-020-00990-y
14. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. 2021. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat Biotechnol 39:578–585. https://doi.org/10.1038/s41587-020-00774-7
15. Steinegger M, Söding J. 2017. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol 35:1026–1028. https://doi.org/10.1038/nbt.3988
16. Mirdita M, Steinegger M, Breitwieser F, Söding J, Levy Karin E. 2021. Fast and sensitive taxonomic assignment to metagenomic contigs. Bioinformatics 37:3029–3031. https://doi.org/10.1093/bioinformatics/btab184
17. Bouras G, Nepal R, Houtak G, Psaltis AJ, Wormald PJ, Vreugde S. 2023. Pharokka: a fast scalable bacteriophage annotation tool. Bioinformatics 39:btac776. https://doi.org/10.1093/bioinformatics/btac776
18. SeemanT. 2020 ABRICATE v1.0.1. Available from: https://github.com/tseemann/abricate
19. Chen LH, Zheng DD, Liu B, Yang J, Jin Q. 2016. VFDB 2016: hierarchical and refined dataset for big data analysis--10 years on. Nucleic Acids Res 44:D694–7. https://doi.org/10.1093/nar/gkv1239
20. Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, Lago BA, Dave BM, Pereira S, Sharma AN, Doshi S, Courtot M, Lo R, Williams LE, Frye JG, Elsayegh T, Sardar D, Westman EL, Pawlowski AC, Johnson TA, Brinkman FSL, Wright GD, McArthur AG. 2017. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res 45:D566–D573. https://doi.org/10.1093/nar/gkw1004
21. Lee I, Ouk Kim Y, Park S-C, Chun J. 2016. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. Int J Syst Evol Microbiol 66:1100–1103. https://doi.org/10.1099/ijsem.0.000760
22. Roux S, Camargo AP, Coutinho FH, Dabdoub SM, Dutilh BE, Nayfach S, Tritt A. 2023. iPHoP: an integrated machine learning framework to maximize host prediction for metagenome-derived viruses of archaea and bacteria. PLoS Biol 21:e3002083. https://doi.org/10.1371/journal.pbio.3002083