bioRxiv preprint doi: https://doi.org/10.1101/858043; this version posted November 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Three complete genome sequences of Penguin megrivirus from ornithogenic soil, Adelie penguin and Weddell seal of Antarctica

Ashutosh Aasdev[a], Anamika Mishra[a], Manoj Nair[b], Satyam D Pawar[a], Chandan K Dubey[a], Sandeep Bhatia[a], Prabir G Dastidar[c], Vijendra Pal Singh[a], Ashwin Ashok Raut[a]*

[a]ICAR-National Institute of High Security Animal Diseases, Bhopal, Madhya Pradesh, India-462022

[b]Wildlife Institute of India, Dehradun, Uttarakhand, India-248001

[c]Ministry of Earth Sciences, Prithvi Bhavan, New Delhi - 110003

Running Head: Penguin megrivirus antarctica soil and wildlife

*Corresponding Author: Ashwin Ashok Raut, ashwin.nihsad@gmail.com.

bioRxiv preprint doi: https://doi.org/10.1101/858043; this version posted November 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Abstract

We report the complete genome sequences of Penguin megriviruses from three different sources from Antarctica namely feces of Adelie penguin, feces of Weddell seal and ornithogenic soil. Phylogenetic analysis indicates the prevalence of very similar viruses in different sources of Antarctic environment. These genome sequences aid to understand the evolution of megriviruses in Antarctic ecology and reveal their place in global megrivirus phylogeny.

bioRxiv preprint doi: https://doi.org/10.1101/858043; this version posted November 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

43 **Announcement**

44 The genus *Megrivirus* belongs to positive sense single stranded RNA virus family *Picornaviridae*.
45 Megrivirus was classified as a genus in 2013 (1). Previously the genus had only one species, the
46 Turkey hepatitis virus, but identification of new viruses has expanded the genus to five species and
47 many unclassified viruses. All megriviruses identified infect aves, and have been identified from
48 chicken to wild birds (2). Penguin megrivirus belongs to species Megrivirus E and was only recently
49 reported in 2017 (3).

50 Samples for this study were collected in February 2018 as part of the 37th Indian Scientific Expedition
51 to Antarctica, Ministry of Earth Sciences, Government of India. RNA extraction from the fecal
52 samples was done on site at Bharati Station, Antarctica. Fresh fecal samples of Adelie penguin and
53 Weddell seal were suspended in PBS and centrifuged at 12,000 RCF for 15 min. The supernatant was
54 filtered (0.45µm) and RNA was extracted using QIAamp Viral RNA mini kit without adding carrier
55 RNA. The RNA samples were dried in GenTegra RNA plates and transported to NIHSAD, Bhopal, India
56 at room temperature. Ornithogenic soil sample transported in cold chain, was processed in the
57 BSL3+ containment lab at NIHSAD. One gram of soil was suspended in 10 mL HBSS and incubated at
58 room temperature for an hour followed by centrifugation at 6,000 RCF for 15 minutes at 4°C. The
59 supernatant was filtered through 0.45µm filter and ultracentrifuged at 180,000 RCF for 4 hours at
60 4°C. The pellet was resuspended in 280 µL HBSS and viral RNA extraction was done using QIAamp
61 Viral RNA Mini Kit.

62 Highthroughput sequence data was generated at Genotypic Technology Private Limited, Bangalore,
63 India by 2 X 150 bp paired end Illumina sequencing on MiSeq platform, using NEBNext® Ultra™
64 Directional RNA Library Prep Kit. The raw reads were quality controlled using FastQC v0.11.3
65 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adapter and low quality bases were
66 removed using Trimmomatic v0.39 (4). BLASTn of the processed reads was performed against a
67 custom viral genome database using local blast v2.8.1+. Reads for individual virus reference

bioRxiv preprint doi: https://doi.org/10.1101/858043; this version posted November 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

sequences were extracted. The reference sequences with high number of reads originating from them were used for mapping the reads using BWA-MEM v0.7.17-r1194-dirty (5). Consensus sequence were evaluated using BCF tools v1.9 (6). PCR confirmation of virus genome was done using primers, PM-F120 5'-TGGAGCTGCCATCACGTGTT-3' and PM-R120 5'-TCCTGACACTGGTCACGTCT-3'. Few of the terminal nucleotides could not be read for the sequences as indicated in Table 1. NCBI ORF finder was used to predict the ORF's. The polyprotein ORF is between nucleotide positions 718 and 9030 and codes for a 2770 amino acid polypeptide. The complete polypeptide sequences of 16 megriviruses was downloaded from NCBI and protein alignment performed using MUSCLE in MEGA7 (7). Neighbour-joining phylogenetic tree was constructed using Poisson model and 1000 bootstrap replicates. The viruses form a monophylectic group. The pairwise amino acid identity amongst the viruses under study is from 99.75% to 99.9%. The 5' UTR secondary structure was predicted using mfold (8). The structure contains elements similar to domain II and III of megrivirus IRES (9, 10). Also, the conserved 20 nucleotides which make the '8' like structure in domain III are present (9, 10). The 3' UTR contains 2 unit A repeats while the AUG rich region following the repeats is absent (10).

In summary we report the complete genome sequences and phylogenetic relation of closely related Penguin megiviruses from three different sources of Antarctica samples. The presence of the virus in fecal samples points to its probable enteric localization. Also, Penguin megrivirus, which is known to infect birds, was identified in Weddell seal feces and there is futher need to investigate its ecological relations and infectivity.

Data Availability

The genome nucleotide sequences have been submitted to GenBank under accession numbers MN453780, MN453781 and MN453782.

bioRxiv preprint doi: https://doi.org/10.1101/858043; this version posted November 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Acknowledgements

Indian Scientific Expeditions to Antarctica (ISEA) by National Center for Polar and Ocean Research, Goa suported by Earth System Science Organization, Ministry of Earth Sciences (ESSO-MoES), Government of India (GoI). Indian Council of Agricultural Research and Director, National Institute of High Security Animal Diseases, Bhopal for deputation of Scientist under ISEA and providing necessary facilities to conduct this work.

bioRxiv preprint doi: https://doi.org/10.1101/858043; this version posted November 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

References

1. Adams M J, King A M Q, Carstens E B. 2013.Ratification vote on taxonomic proposalsto the International Committee on Taxonomy of Viruses. *Archives of Virology* **158**:2023-2030.

2. Boros A, Pankovics P, MáticsR, Adonyi A, Bolba N, Phan T G, Delwart E, Reuter G. 2017. Genome characterization of a novel megrivirus-related avian picornavirus from a carnivorous wild bird, western marsh harrier (*Circus aeruginosus*). *Archives of Virology* **162**:2781-2789.

3. Zell R. 2017. Picornaviridae — the ever growing virus family. *Archives of Virology* **163**:299-317.

4. Bolger A M, Lohse M, Usadel B. 2014. Genome analysis Trimmomatic : a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**:2114–2120.

5. Li, H. 2013. Aligning sequence reads , clone sequences and assembly contigs with BWA-MEM. *arXiv* 1303.

6. Li H. 2011. A statistical framework for SNP calling , mutation discovery , association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**:2987–2993.

7. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* **33**:1870–4.

bioRxiv preprint doi: https://doi.org/10.1101/858043; this version posted November 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

138 **8.** Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction.
139 *Nucleic Acids Research* **31**:3406–3415.
140
141 **9.** Wang F, Liang T, Liu N, Ning K, Yu K, Zhang D. 2017. Genetic characterization of two novel
142 megriviruses in geese. *Journal of General Virology* **98**:607-611.
143
144 **10.** Boros A, Pankovics P, Knowles N J, Nemes C, Delwart E, Reuter G. 2014. Comparative
145 Complete Genome Analysis of Chicken and Turkey Megriviruses (Family *Picornaviridae*):
146 Long 3' Untranslated Regions with a Potential Second Open Reading Frame and Evidence for
147 Possible Recombination. *J Virol* **88**:6434-6443.
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163

bioRxiv preprint doi: https://doi.org/10.1101/858043; this version posted November 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Table 1: Sampling location and sequencing metrics of the Penguin megriviruses from Antarctica

Sample	Accession Numbers	Location	Total number of reads	Number of mapped reads	Avearage coverage	Percentage genome obtained	Number of missing terminal nucleotides	G+C content (%)
Ornitho-genic soil	MN453780	Svenner Islands, Prydz Bay (-69.03, 76.83)	176,926,262	4,273,385	62,167	99.93	4	42.87
Adelie penguin feaces	MN453781	Svenner Islands, Prydz Bay (-69.03, 76.83)	144,867,534	11,520	169	99.88	13	43.03
Weddell seal feaces	MN453782	Larsemann Hills, Prydz Bay (-69.58, 78.55)	133,147,552	3,085	44	99.92	8	43.04