MPoxVR: A comprehensive genomic resource for monkeypox virus variant surveillance

Yingke Ma,1,3 Meili Chen,1,3 Yiming Bao,1,2,,* Shuhui Song,1,* and MPoxVR Team

1National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences and China National Center for Bioinformation, Beijing 100101, China
2University of Chinese Academy of Sciences, Beijing 100049, China
3These authors contributed equally
*Correspondence: baoym@big.ac.cn (Y.B.); songshh@big.ac.cn (S.S.)
Received: June 18, 2022; Accepted: July 28, 2022; Published Online: August 1, 2022; https://doi.org/10.1016/j.xinn.2022.100296
© 2022 The Authors. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Citation: Ma Y., Chen M., Bao Y., et al., and MPoxVR Team (2022). MPoxVR: A comprehensive genomic resource for monkeypox virus variant surveillance. The Innovation 3(5). 100296.

Monkeypox is a viral zoonotic disease endemic in Central and West Africa. Since January 1, 2022, 3413 laboratory-confirmed monkeypox cases and one death have been reported from 50 countries/territories in five WHO regions (as of June 22, 2022; https://www.who.int/emergencies/disease-outbreak-news/item/2022-DON396), and 1310 new cases and eight new countries have been reported in the past week. Genomic epidemiology is vital to determine the similarity between viruses and suggest possible links between cases, origins of infection, and transmission dynamics when combined with epidemiological information. However, one of the priority evidence gaps relating to the monkeypox outbreak is genome sequencing and in-host variation analysis.1 So, timely sharing both raw sequence data and consensus genomic data are useful to public health investigators and academic partners undertaking related studies. Several databases have created pox- or monkeypox-related resources. These include Global Initiative on Sharing All Influenza Data (GISAID; https://www.epicov.org/epi3/),

Figure 1. Statistical distribution of monkeypox virus genome sequence and variation (A) Distribution of monkeypox virus genome sequence. (B) Statistical distribution of SNP and affected amino acid variants in monkeypox virus complete genome sequences since the outbreak in May 2022.
National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/labs/virus/vssi/), NextStrain (https://nextstrain.org/monkeypox/hmpxv1), Virus Pathogen Resource (ViPR; https://viprbrc.org/brc/home.spg?decorator=px), and Viralzone (https://viralzone.expasy.org/9957). Most of these resources copy genomic sequences and metadata from GenBank without value-added curation. In addition, they lack sequence variation analysis, tools, or publications specific to monkeypox virus.

To help the scientific community to rapidly access the monkeypox genomic information, China National Center for Bioinformation (CNGB)-National Genomics Data Center (NGDC) created Monkeypox Virus Resource (MPoxVR; https://ngdc.cnbc.ac.cn/gwh/poxvirus/), an open-access information resource on the monkeypox virus (MPXV). MPoxVR features integration of MPXV genome sequences and value-added metadata and offers personalized data retrieval and download services. As of July 26, 2022, MPoxVR has integrated 767 MPXV sequences (Figure 1A) from 17 different hosts from NCBI. Among them, 403 (53%) have a complete genome that covers all protein coding genes. Strikingly, 6% of the sequences are missing critical metadata information, such as virus host, sample collection date, or sampling geographic location, which are important for identifying the transmission networks and understanding the mode of transmission. Even by the monkeypox outbreak in May 2022, the specific sampling date is incomplete for 206 (51%) complete sequences. By manual curation from related publications, the missing or incomplete metadata for most complete sequences in MPoxVR were added except for three sequences, which significantly improved the quality and availability of genome data. MPoxVR performs comprehensive genomic variation analyses and provides all identified variants and detailed statistics for each virus isolate based on sequence alignment against the first public MPXV sequence from the 2022 outbreak (ON563414.3) and congregates the functional annotation and population frequency for each variant. Based on variants statistic in MPoxVR, a small difference ranging from 0 to 11 SNPs between sequences of the monkeypox outbreak in May 2022 was observed (Figure 1B), and individual sequences have abnormally high number of variants (e.g., ON675438.1), which may be introduced by sequencing errors. More importantly, most mutations are new in 2022, and some mutations in genes involved in immune evasion, host range, cell proliferation (e.g., A45L, C7L, D7L, D10L), drug resistance (e.g., L3R, L6R, and A25R), and vaccine development (e.g., A25R) currently bear low mutation frequency, but close monitoring is still warranted. No new mutation was found in the primer regions of nucleic acid detection. When comparing the current MPXV strains in the UK with the one in 2018 (GenBank: MT903345), a small subset of mutations was also found in genes involved in virus transmission, virulence, or interaction with antiviral drugs. Recently, mutations among sequences in two distinct lineages detected in the United States in 2021–2022 showed an extreme preference for GA-to-AA mutations, indicative of APOBEC3 cytosine deaminase activity that was shared among West African MPXVs since 2017 but was absent from Congo Basin lineages. In all, accurate identification of variants between sequences is important for functional and evolutionary analysis.

Of particular note, MPoxVR has established four commonly used online tools, including BLAST, genome annotation, variant identification, and variant annotation, which can be freely accessed at https://ngdc.cnbc.ac.cn/gwh/poxvirus/tools. Genome annotation is essential for performing sequence comparison and evolutionary analysis. To accurately annotate the gene structure of MPXVs, an online genome annotation tool was developed based on VAPiD, which enables accurate identification of coding sequences and protein sequences in about 30 s and provides annotation files to facilitate virus genome submissions to NCBI-NGDC Genome Warehouse and NCBI GenBank. By comparing them against the MPXV reference sequence (GenBank: ON563414.3) with MUMmer 3.8.31, sequence variants could be identified using in-house developed programs (the scripts are shared at the https://ngdc.cnbc.ac.cn/gwh/poxvirus/tools), and the mutation effects were determined using Ensembl Variant Effect Predictor in about 10 s. Besides, by linking with relevant databases in NCBI-NGDC, MPoxVR offers data submission services for raw sequence reads and assembled genomes and data sharing with NCBI. MPoxVR also provides a full collection of articles on MPXV, including published papers from PubMed as well as preprints from services such as bioRxiv and medRxiv through Europe PMC.

Compared with other poxvirus-related data resources, MPoxVR features timely integration of publicly released genome sequences and curation of value-added metadata, which are particularly important for identifying the transmission networks. MPoxVR performs comprehensive genomic variation analyses to provide all identified variants and detailed statistics for each virus isolate and congregate functional annotation and the population frequency for each variant. MPoxVR also features a full collection of literature on MPXV. MPoxVR is working on developing an MPXV sequence and variant monitoring and tracking system to better serve the monkeypox research community. We call on worldwide collaborations on data sharing and welcome any feedback for further improvement.

REFERENCES

1. Charlotte, A., Carolina, A., Sooria, B., et al. (2022). Investigation into Monkeypox Outbreak in England: Technical Briefing 1 (UK Health Security Agency). https://www.gov.uk/government/publications/monkeypox-outbreak-technical-briefings/investigation-into-monkeypox-outbreak-in-england-technical-briefing-1part-3-genomics.
2. Gigante, C.M, Korber, B, Seabolt, M.H., et al. (2022). Multiple lineages of Monkeypox virus detected in the United States, 2021–2022. Preprint at bioRxiv https://doi.org/10.1101/2022.06.10.495526.
3. Shean, R.C., Makhous, N., Stoddard, G.D., et al. (2019). VAPiD: a lightweight cross-platform viral annotation pipeline and identification tool to facilitate virus genome submissions to NCBI GenBank. BMC Biomed. 20, 48.
4. Kurtz, S., Phillippy, A., Delcher, A.L, et al. (2004). Versatile and open software for comparing large genomes. Genome Biol. 5, R12.
5. McLaren, W., Gil, L., Hunt, S.E., et al. (2016). The Ensembl variant effect predictor. Genome Biol. 17, 122.

ACKNOWLEDGMENTS

This research was funded by grant from the Youth Innovation Promotion Association of CAS (2021038 to S.S.). We are grateful to other members of the MPoxVR team, whose names are not included as authors due to the journal’s policy (Cuiping Li, Shuai Jiang, Dong Zou, Zheng Gong, Xuetong Zhao, Yangqing Wang, Junwei Zhu, Zhang Zhang, Wemm- ing Zhao, and Yongbiao Xue).

DECLARATION OF INTERESTS

The authors declare no competing interests.