



Genome Sequences of Two New Pandoravirus Strains Isolated from Brazil and France

 Djamal Brahim Belhaouari,^{a,b}  Sihem Hannat,^{a,b}  Rodrigo Rodrigues,^c Sarah Aherfi,^{a,b}  Bernard La Scola,^{a,b}  Julien Andreani^{a,b,*}

^aAix-Marseille Université, Institut de Recherche pour le Développement (IRD), Microbes, Evolution, Phylogeny and Infection (MEPHI), Marseille, France

^bInstitut Hospitalo-Universitaire (IHU)—Méditerranée Infection, Marseille, France

^cLaboratório de Vírus, Departamento de Microbiologia, Instituto de Ciências Biológicas Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

ABSTRACT Pandoraviruses are giant viruses of amoebas with a wide range of genome sizes (1.5 to 2.5 Mbp) and 1- μ m ovoid viral particles. Here, we report the isolation, genome sequencing, and annotation of two new strains from the proposed family Pandoraviridae: *Pandoravirus belohorizontensis* and *Pandoravirus aubagnensis*.

Pandoraviruses are giant viruses as large as bacteria and have more complex genomes than some eukaryotic organisms (1). Several pandoraviruses have been described using coculture on *Acanthamoeba castellanii* (2–7). Here, we report the complete genome sequences of two novel strains: *Pandoravirus belohorizontensis*, isolated from soil samples collected from the city of Belo Horizonte (–19.923249777699425, –43.93308441843789), and *Pandoravirus aubagnensis*, from water collected in the south of France (Mounoi Cavern, also called “Manon des Sources”; 43.274642, 5.777029). The two viruses were isolated following a procedure previously described by Khalil et al. (8). Briefly, samples were cocultured on *Acanthamoeba castellanii*. They were characterized using flow cytometry and electron microscopy and then were produced and purified for genome sequencing. Viral DNA was extracted using an EZ1 Advanced XL automated system (Qiagen, France). DNA paired-end libraries (2 × 250-bp) were constructed with 1 ng of each genome as input using the Nextera XT DNA kit (Illumina, Inc., San Diego, USA) and sequenced on the Illumina MiSeq instrument. The reads were then trimmed and filtered using Trimmomatic (9). The *P. belohorizontensis* genome was assembled using CLC Genomics Workbench v7.5.2. The genome was finished using MUMmer v3.0 with default parameters (10), followed by a genome scaffolder using a graph-based approach (11). The genome of *P. aubagnensis* was assembled using SPAdes (12) and joined into a single scaffold using scaffold_builder (13). The genome termini were verified using Mauve software (14) and by a BLASTn search of both genomes against the nonredundant nucleotide (nr/nt) database (15). The analysis of both genomes followed the same procedure with default parameters. tRNAs were predicted using tRNAscan-SE (16) and ARAGORN (17) software. Gene predictions were performed using GeneMarkS (18). Predicted proteins over 99 amino acids long were considered for further analysis. The predicted proteins were investigated for putative functions and domains using BLASTp searches (E values, <1E-03) against the nonredundant protein database and the Pfam protein families database (19) and using Delta-BLAST (20). Phylogenetic analysis was based on the DNA polymerase subunit B gene. Amino acid sequences were aligned using Muscle (21). The maximum likelihood method was used for tree construction on MEGA7 (22) with the Jones-Taylor-Thornton model for amino acid substitution. The collection and analysis of genetic data were partially or fully registered under SisGen permit number AC31840 and SISBIO license numbers 33326, 34293, and 80252 (Brazil).

The *P. belohorizontensis* genome was assembled into a single scaffold of 1,701,725 bp (average coverage, 223×) with 19 gaps of unknown length and a G+C content of 63.67%. The *P. belohorizontensis* genome was predicted to encode 1,059 proteins (mean size \pm SD,

Editor Simon Roux, DOE Joint Genome Institute

Copyright © 2022 Brahim Belhaouari et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Bernard La Scola, bernard.la-scola@univ-amu.fr, or Julien Andreani, JAndreani@chu-grenoble.fr.

*Present address: Julien Andreani, Université Grenoble Alpes, Laboratoire de Virologie, Centre Hospitalier Universitaire (CHU) Grenoble Alpes, Grenoble, France.

The authors declare no conflict of interest.

Received 25 March 2022

Accepted 2 June 2022

Published 22 June 2022

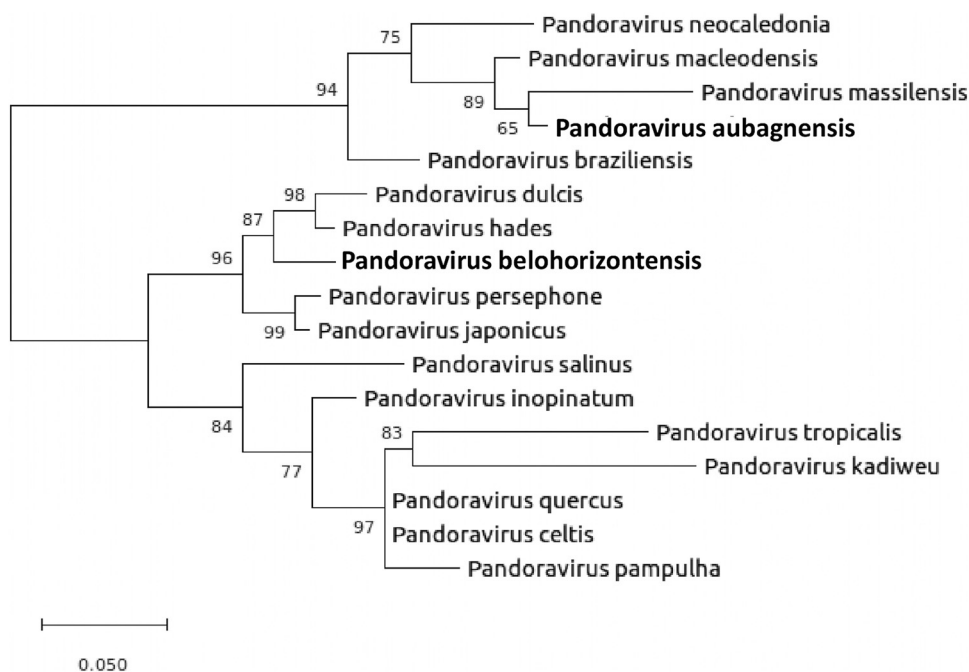


FIG 1 Phylogenetic reconstruction based on amino acid sequences of the DNA polymerase B subunit of Pandoravirus. The phylogenetic tree was built using the maximum likelihood model with 1,000 bootstrap replicates. The Pandoravirus kadiweu, P. tropicalis, P. pampulha, P. hades, and P. persephone sequences are partial predicted proteins (scale bar indicates 0.05 substitutions/site).

363 ± 248 amino acids). Of these, 883 (83.4%) have a homolog in the nr/nt database, and 176 (16.6%) are ORFans (open reading frames [ORFs] with no significant homolog in the nr/nt database).

The assembly of the *P. aubagnensis* genome provided a single scaffold of 1,816,783 bp (average coverage, 198×) with 6 gaps of estimated length and a G+C content of 58.02%. A total of 1,217 proteins were predicted (mean size ± SD, 345 ± 244 amino acids). Of these, 907 (74.6%) have a homolog in the nr/nt database, and 309 (25.4%) are ORFans. tRNA prediction showed that both genomes encode a single proline tRNA.

Phylogenetic analysis revealed that the two isolates were different from each other and clustered with previously described Pandoravirus lineages (Fig. 1).

Data availability. The genome sequences of Pandoravirus belohorizontensis and Pandoravirus aubagnensis have been deposited at NCBI GenBank under the accession numbers [MZ420562](https://doi.org/10.1093/nar/nzab001) and [MZ420563](https://doi.org/10.1093/nar/nzab002) and the annotation and SRA data under the SRA accession numbers [SRR17644538](https://doi.org/10.1093/bioinformatics/btab001) and [SRR17635305](https://doi.org/10.1093/bioinformatics/btab002), respectively.

ACKNOWLEDGMENTS

We thank A. Levasseur for his help in assembling the genome of Pandoravirus belohorizontensis.

This work was supported by the French Government under the program “Investments for the Future,” managed by the National Agency for Research (ANR), Méditerranée-Infection (10-IAHU-03). It was also supported by Région Provence-Alpes-Côte d’Azur and the European funding organization FEDER PRIMMI.

No potential conflicts of interest or financial disclosures are reported for any authors.

REFERENCES

- Philippe N, Legendre M, Doutré G, Couté Y, Poirot O, Lescot M, Arslan D, Seltzer V, Bertaux L, Bruley C, Garin J, Claverie J-M, Abergel C. 2013. Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. *Science* 341:281–286. <https://doi.org/10.1126/science.1239181>.
- Aherfi S, Andreani J, Baptiste E, Oumessoum A, Dornas FP, Andrade ACDS, Chabriere E, Abrahao J, Levasseur A, Raoult D, La Scola B, Colson P. 2018. A large open pangenome and a small core genome for giant pandoraviruses. *Front Microbiol* 9:1486. <https://doi.org/10.3389/fmicb.2018.01486>.
- Legendre M, Alempic J-M, Philippe N, Lartigue A, Jeudy S, Poirot O, Ta NT, Nin S, Couté Y, Abergel C, Claverie J-M. 2019. Pandoravirus Celtis illustrates

- the microevolution processes at work in the giant Pandoraviridae genomes. *Front Microbiol* 10:430. <https://doi.org/10.3389/fmicb.2019.00430>.
4. Andrade ACDSP, Boratto PVDM, Rodrigues RAL, Bastos TM, Azevedo BL, Dornas FP, Oliveira DB, Drummond BP, Kroon EG, Abrahão JS. 2019. New isolates of pandoraviruses: contribution to the study of replication cycle steps. *J Virol* 93:e01942-18. <https://doi.org/10.1128/JVI.01942-18>.
 5. Akashi M, Takemura M. 2019. Co-isolation and characterization of two pandoraviruses and a mimivirus from a riverbank in Japan. *Viruses* 11: 1123. <https://doi.org/10.3390/v11121123>.
 6. Hosokawa N, Takahashi H, Aoki K, Takemura M. 2021. Draft genome sequence of Pandoravirus japonicus isolated from the Sabaishi River, Niigata, Japan. *Microbiol Resour Announc* 10:e00365-21. <https://doi.org/10.1128/MRA.00365-21>.
 7. Legendre M, Fabre E, Poirot O, Jeudy S, Lartigue A, Alempic J-M, Beucher L, Philippe N, Bertaux L, Christo-Foroux E, Labadie K, Couté Y, Abergel C, Claverie J-M. 2018. Diversity and evolution of the emerging Pandoraviridae family. *Nat Commun* 9:2285. <https://doi.org/10.1038/s41467-018-04698-4>.
 8. Bou Khalil JY, Andreani J, Raoult D, La Scola B. 2016. A rapid strategy for the isolation of new faustoviruses from environmental samples using *Vermamoeba vermiformis*. *J Vis Exp*:54104. <https://doi.org/10.3791/54104>.
 9. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
 10. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. *Genome Biol* 5:R12. <https://doi.org/10.1186/gb-2004-5-2-r12>.
 11. Bosi E, Donati B, Galardini M, Brunetti S, Sagot M-F, Lió P, Crescenzi P, Fani R, Fondi M. 2015. MeDuSa: a multi-draft based scaffold. *Bioinformatics* 31:2443–2451. <https://doi.org/10.1093/bioinformatics/btv171>.
 12. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
 13. Silva GG, Dutilh BE, Matthews TD, Elkins K, Schmieder R, Dinsdale EA, Edwards RA. 2013. Combining de novo and reference-guided assembly with scaffold_builder. *Source Code Biol Med* 8:23. <https://doi.org/10.1186/1751-0473-8-23>.
 14. Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14:1394–1403. <https://doi.org/10.1101/gr.2289704>.
 15. Pruitt KD, Tatusova T, Maglott DR. 2005. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 33:D501–D504. <https://doi.org/10.1093/nar/gki025>.
 16. Schattner P, Brooks AN, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS Web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res* 33:W686–W689. <https://doi.org/10.1093/nar/gki366>.
 17. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* 32:11–16. <https://doi.org/10.1093/nar/gkh152>.
 18. Besemer J, Lomsadze A, Borodovsky M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res* 29: 2607–2618. <https://doi.org/10.1093/nar/29.12.2607>.
 19. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* 44:D279–D285. <https://doi.org/10.1093/nar/gkv1344>.
 20. Marchler-Bauer A, Bryant SH. 2004. CD-Search: protein domain annotations on the fly. *Nucleic Acids Res* 32:W327–W331. <https://doi.org/10.1093/nar/gkh454>.
 21. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <https://doi.org/10.1093/nar/gkh340>.
 22. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 35:1547–1549. <https://doi.org/10.1093/molbev/msy096>.