

# Complete Genome Sequence of a Novel Marafivirus Infecting Pearl Millet in Burkina Faso

**Essowè Palanga**

ITRA: Institut Togolais de Recherche Agronomique

**Ezechiël Bionimian Tibiri**

INERA: Institut de l'Environnement et de Recherches Agricoles

**Martine Bangratz**

IRD: Institut de recherche pour le developpement

**Denis Filloux**

CIRAD

**Charlotte Julian**

CIRAD

**Agnès Pinel-Galzi**

IRD: Institut de recherche pour le developpement

**Moustapha Koala**

INERA: Institut de l'Environnement et de Recherches Agricoles

**James Bouma Néya**

INERA: Institut de l'Environnement et de Recherches Agricoles

**Christophe Brugidou**

IRD: Institut de recherche pour le developpement

**Fidèle Tiendrébéogo**

INERA: Institut de l'Environnement et de Recherches Agricoles

**Philippe Roumagnac**

CIRAD

**Eugénie Hébrard** (✉ [eugenie.hebrard@ird.fr](mailto:eugenie.hebrard@ird.fr))

Institut de Recherche pour le Developpement <https://orcid.org/0000-0002-2128-1302>

---

## Research Article

**Keywords:** Tymoviridae, Marafivirus, PGMV, Pearl millet, Burkina Faso

**Posted Date:** July 20th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-720372/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Archives of Virology on October 25th, 2021.

See the published version at <https://doi.org/10.1007/s00705-021-05262-0>.

# Abstract

Pearl millet (*Pennisetum glaucum* (L.) R. Br.) is a staple food that is widely cultivated in sub-Saharan Africa. In August 2018, a survey was conducted in the main producing regions of Burkina Faso and leaf samples were analyzed using virion-associated nucleic acid (VANA)-based metagenomic approach and Illumina sequencing. A new virus, tentatively named “pennisetum glaucum marafivirus” (PGMV), was detected and its complete nucleotide sequence of 6364 nucleotides was determined. The sequence contains a large open reading frame (ORF) encoding a polyprotein of 224.2 kDa with five domains (methyltransferase, papain-like protease, helicase, RNA-dependent RNA polymerase and coat proteins), typical of members of the marafiviruses. Additionally, a characteristic conserved marafibox domain was detected in the genome. The nucleotide sequence of the PGMV complete genome shares 68.5% identity with *Sorghum bicolor marafivirus* and its coat protein shares 58.5% identity with *Oat blue dwarf virus*. Phylogenetic analyses confirmed that the pearl millet virus is unambiguously grouped with members of the genus *Marafivirus* in the family *Tymoviridae*. This is the first report on the occurrence of a marafivirus on pearl millet.

## Background

Pearl millet (*Pennisetum glaucum* (L.) R. Br.) is a staple food in Africa. Faced with climate change, this dryland cereal would play a major role in food security in the next future. However, crop intensification often leads to the emergence of viral diseases. In the 1990-2000s, almost ten viruses have been reported on pearl millet: *Johnsongrass mosaic virus*, *Maize dwarf mosaic virus* and *Sugarcane mosaic virus* (*Potyvirus*, *Potyviridae*), *Wheat streak mosaic virus* (*Tritimovirus*, *Potyviridae*), *Maize streak virus* (*Mastrevirus*, *Geminiviridae*), *Panicum mosaic virus* (*Panicovirus*, *Tombusviridae*) and the associated *Panicum mosaic satellite virus*, *Indian peanut clump virus* (*Pecluvirus*, *Virgaviridae*) and *Rice black streaked dwarf virus* (*Fijivirus* *Reoviridae*), [1, 2], but current epidemiological data and yield losses are poorly documented.

Here, we identified and characterized a full genome sequence of a novel marafivirus infecting pearl millet in Burkina Faso. In August 2018, pearl millet leaves showing small brown dots were collected in a farmer's field at Nouna, Diébégou districts, Burkina Faso. Calcium chloride dihydrate (CaCl<sub>2</sub>) was used for leaves drying. After collecting leaves, we first placed them in absorbent paper and introduced them into a 50 ml falcon tube containing about 20g CaCl<sub>2</sub>. Tubes were then immediately stored at 4°C for 48 hours. Finally the samples were stored at room temperature for further analysis using the virion-associated nucleic acid (VANA)-based metagenomic approach as previously described [3]. Purified DNAs libraries were sequenced using Illumina HiSeq platform in 2x150 bp paired-end reads (Genewiz, South Plainfield, NJ, USA). Illumina sequences were analyzed following a bioinformatic pipeline as previously described [3]. Contigs produced by read assembly were utilized as queries to perform BLASTn and BLASTx searches [4] using the NCBI database (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). These BLAST searches revealed the presence of several Marafivirus-like contigs (Supplementary Table 1). Total RNAs were further extracted using the GeneJET Plant RNA Purification Kit (Thermo Fisher Scientific, Waltham, MA USA).

cDNAs were constructed with reverse primer oligodT and MMLV-RT (200 u/μl) enzyme kit (Promega, Fitchburg Madison/USA) following recommendations. Internal and external parts of the genome were reconstituted and validated using primer pairs designed from Illumina sequences: Marafi-1662F/Marafi-2175R (~ 514 bp); Marafi-5974F/Marafi-polyT (~ 315 bp) and Marafi-1F/Marafi-411R (~ 411 bp) in independent PCR amplifications using Takara (5 u/μl) enzyme (Supplementary table 1). To obtain the 3' end sequence, a 3'RACE PCR was performed as described in [5] using primer combination Marafi-5974F/M4T/M4. Next amplicon was sequenced using primers Marafi-6110F, Marafi-6237F and M4 (Supplementary table 1). The 5'RACE PCR was performed using the Nanopore MinION cDNA-PCR sequencing method according to Oxford Nanopore Technologies instructions (version: PCS\_9085\_v109\_revK\_14Aug2019). cDNAs were constructed using total RNAs and a reverse primer Marafi-411R in a first step and a Nanopore switching primer SSP in a second step (Supplementary table 1). Two independent PCRs were performed, adopting primer pairs PR2/Marafi-159R (~ 159 bp) and PR2/Marafi-198R (~ 200 bp) (Supplementary table 1). PCR products were directly sequenced with Sanger method using primers Marafi-159R and Marafi-198R, respectively. All Sanger sequences were edited and assembled with Illumina contigs using Geneious v5.5.9 software (<https://www.geneious.com>) to reconstitute the marafivirus whole genome. Open reading frames (ORFs) were detected using the ORF Finder NCBI analysis tool (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>). Genome-wide pairwise identity comparison of the pearl millet-derived marafivirus and representative sequences of the *Tymoviridae* family were performed using SDT v.1.2 [6]. Complete genome sequences and coat protein sequences were aligned using MUSCLE 3.7 with default settings [7] and phylogenetic relationships were performed using PhyML 3.1 [8] implemented in MEGAX software [9] using GTR + G + I model for complete genome sequences and LG + G + F model for coat protein sequences. One thousand bootstrap replicates were implemented as a test for the support of branches.

The complete genome of the novel pearl millet virus was 6364 nt long without the polyA tail. A large ORF was detected from 156 to 6218 nt (6063 nt) which encodes a 2020 aa-long polyprotein (224.2 kDa) (Fig. 1A). Several domains were identified including methyltransferase (MTR, 36–317 aa), papain-like protease (PRO, 722–817 aa), helicase (HEL, 913–1144 aa), RNA-dependent RNA polymerase (RdRp, 1448–1664 aa) and two forms of the same coat protein (CP, 1847–2019 aa) (Fig. 1A). The novel genome also contains a conserved “marafibox”, a 16-nt consensus sequence present in all known marafiviruses, from 5463 to 5478 nt (Fig. 1A and 1B). This domain is different from other *Poaceae* marafiboxes at its fifteenth nucleotide with “**C**” but different from “tymobox” in four nucleotide sites (Fig. 1, B) [10, 11]. Furthermore, the conserved adenine A<sub>(5488 nt)</sub> indicating the initiation site of the CPs sgRNA transcription was detected at ~ 10 nt downstream the marafibox in the conserved nt CAA<sub>(5487–5489nt)</sub> (Fig. 1B) [11–13]. Two methionine codons were detected downstream the marafibox at positions 5523 and 5631, they could be the putative start codons of a minor 25kDa coat protein (CP<sub>1</sub>: 1790–2020 aa, 24.5 kDa) and a major 21kDa coat protein (CP<sub>2</sub>: 1826–2020 aa, 20.8 kDa). Nucleotide pairwise identities showed that the complete genome shared 68.5% sequence identity (Supplementary table 3) with sorghum bicolor marafivirus USA (MN100128) [14]. The CP protein sequence shared 58.5% identity with oat blue dwarf virus (OBDV, U87832) and 56% identity with sorghum bicolor marafivirus USA (MN100128)

(Supplementary table 3). Phylogenetic trees confirmed that both the complete genome and the CP protein sequences of the novel virus isolated from pearl millet grouped with *Marafivirus* species (Fig. 2, A and B). According to the ninth report of International Committee on Taxonomy of Viruses (ICTV), new virus species might be created in the genus *Marafivirus* if the overall sequence identity is less than 80% and the capsid protein sequences are less than 90% identical [10]. Our results are in agreement with this demarcation criterion. In conclusion, the new virus, tentatively named pennisetum glaucum marafivirus (PGMV) could be considered as a putative new species that belongs to the genus *Marafivirus*, family *Tymoviridae*.

## Declarations

### Acknowledgments

E. Palanga holds a post-doctoral fellowship funded by the program Make Our Planet Again (MOPGA). This project was supported by the CGIAR Research Program on Grain Legumes and Dryland Cereals (CRP GLDC). We thank the farmers from Diébégou for their kind collaboration. This work was performed thanks to the facilities of the “International joint Laboratory LMI PathoBios: Observatory of plant pathogens in West Africa: biodiversity and biosafety” ([www.pathobios.com](http://www.pathobios.com); [twitter.com/PathoBios](https://twitter.com/PathoBios)). We are grateful to the CIRAD Southgreen and INRAE MIGALE bioinformatics facilities for providing computing resources.

### Compliance with Ethical Standards

In every case, we obtained permission from the farmers to work in their fields, and the management of the entire project followed the guidelines of the Nagoya protocol regarding access and benefit-sharing.

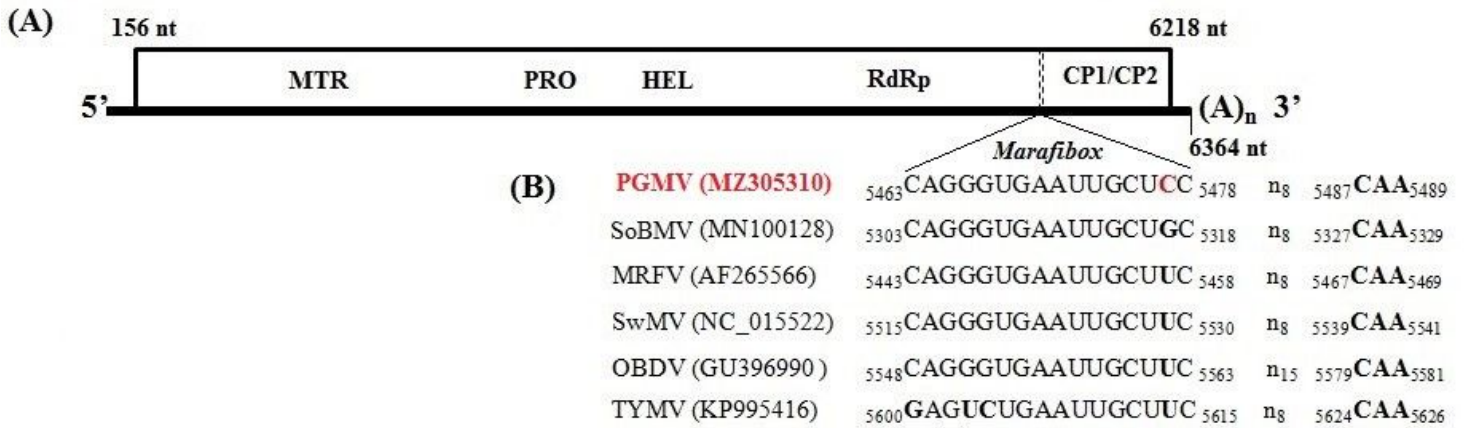
## References

1. Fauquet C, Thouvenel JC (1987). Maladies virales des plantes en Côte d'Ivoire = Plant viral diseases in the Ivory Coast. ORSTOM, Paris, Initiations - Documentations Techniques N°46, 344 p. ISBN 2-7099-0846-8
2. Wilson JP (1999) Pearl millet diseases: A compilation of Information on the Known Pathogens of Pearl Millet Pennisetum glaucum (L.) R. Br. U.S. Department of Agriculture, Agricultural Research Service, Agriculture Handbook N°176.
3. François S, Filloux D, Fernandez E, et al (2018) Viral Metagenomics Approaches for High-Resolution Screening of Multiplexed Arthropod and Plant Viral Communities. *Methods Mol Biol* 1746:77–95. [https://doi.org/doi:10.1007/978-1-4939-7683-6\\_7](https://doi.org/doi:10.1007/978-1-4939-7683-6_7)
4. Altschul SF, Gish W, Miller W, et al (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
5. Palanga E, Martin DP, Galzi S, et al (2017) Complete genome sequences of cowpea polerovirus 1 and cowpea polerovirus 2 infecting cowpea plants in Burkina Faso. *Arch Virol* 162:2149–2152.

<https://doi.org/10.1007/s00705-017-3327-z>

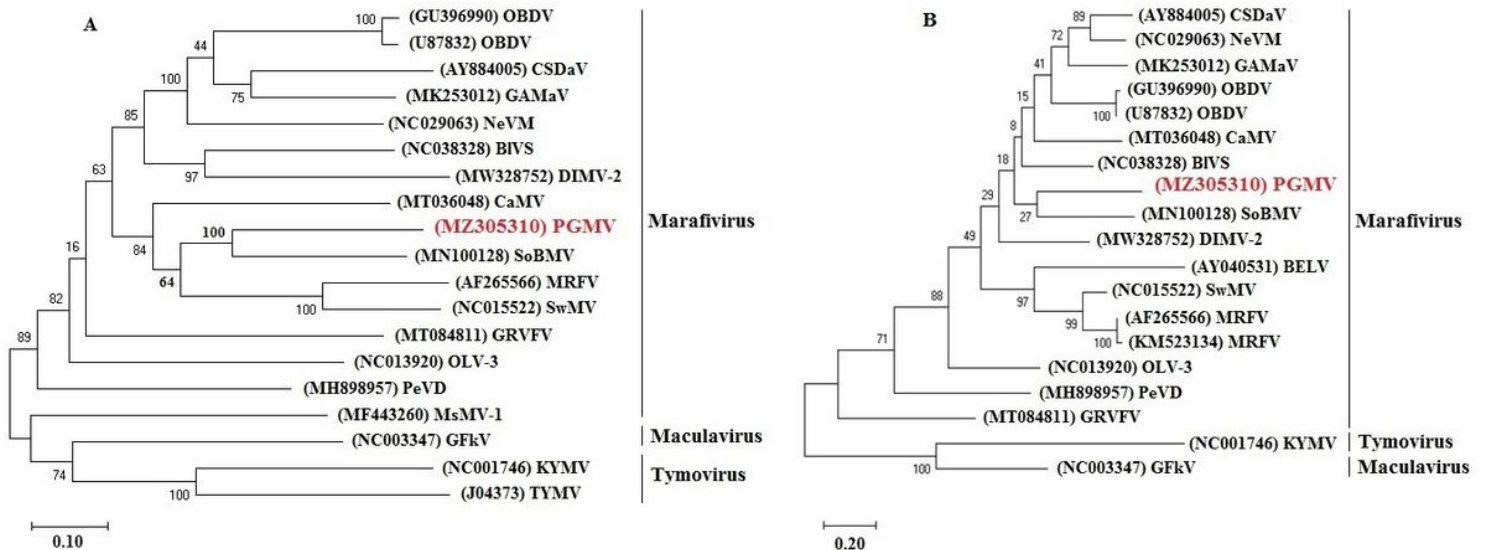
6. Muhire BM, Varsani A, Martin DP (2014) SDT: A virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS One* 9:e108277. <https://doi.org/10.1371/journal.pone.0108277>
7. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <https://doi.org/10.1093/nar/gkh340>
8. Guindon S, Dufayard JF, Lefort V, et al (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol* 59:307–321. <https://doi.org/10.1093/sysbio/syq010>
9. Kumar S, Stecher G, Li M, et al (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 35:1547–1549. <https://doi.org/10.1093/molbev/msy096>
10. Dreher, T.W., Edwards, M.C., Gibbs, A.J., Haenni, A.-L., Hammond, R.W., Jupin, I., Koenig, R., Sabanadzovic, S. and Martelli, G.P. (2012) Family Tymoviridae. Pages 944-952 in: *Virus Taxonomy. Ninth Report of the International committee on Taxonomy of Viruses*. King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ eds. Elsevier Academic Press, Oxford. <https://doi.org/10.1016/B978-0-12-384684-6.00136-1>
11. Edwards MC, Weiland JJ (2014) Coat protein expression strategy of oat blue dwarf virus. *Virology* 450–451:290–296. <https://doi.org/10.1016/j.virol.2013.12.018>
12. Maccheroni W, Alegria MC, Greggio CC, et al (2005) Identification and Genomic Characterization of a New Virus (Tymoviridae Family) Associated with Citrus Sudden Death Disease. *J Virol* 79:3028–3037. <https://doi.org/10.1128/jvi.79.5.3028-3037.2005>
13. Hammond RW, Ramirez P (2001) Molecular characterization of the genome of Maize rayado fino virus, the type member of the genus Marafivirus. *Virology* 282:338–347. <https://doi.org/10.1006/viro.2001.0859>
14. Boukari W, Mollov D, Wei C, et al (2021) Screening for sugarcane yellow leaf virus in sorghum in Florida revealed its occurrence in mixed infections with sugarcane mosaic virus and a new marafivirus. *Crop Prot* 139:105373. <https://doi.org/10.1016/j.cropro.2020.105373>

## Figures



**Figure 1**

PGMV genome organization: (A), marafiviruses large ORF with five typical conserved protein domains: methyltransferase (MTR); papain-like protease (PRO); helicase (HEL); RNA-dependent RNA polymerase (RdRp); coat proteins (CP1: minor or larger and CP2: major or smaller) and marafibox represented by dashed lines. (B), sequence alignment of the marafibox and the initiation site (CAA) position of the CPs sgRNA transcription. A single mutation was detected within the PGMV sequence compared to other Gramineae marafiviruses or to the tymobox of turnip yellow mosaic virus (TYMV, Tymovirus).



**Figure 2**

A) Maximum-likelihood phylogenetic tree of complete genome nucleotide sequences of PGMV and representative members of the Tymoviridae family with 1000 bootstrap replicates. B) Maximum-likelihood phylogenetic tree of coat protein amino acid sequences of PGMV and representative members of the Tymoviridae family with 1000 bootstrap replicates.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [MZ305310.docx](#)
- [Marafivirussupplv210713.docx](#)