

METHODOLOGY

Open Access



A ribodepletion and tagging protocol to multiplex samples for RNA-seq based virus detection: application to the cassava virome

Daniel H. Otron^{1,2,3}, Justin S. Pita^{1,3}, Murielle Hoareau², Fidèle Tiendrébéogo¹, Jean-Michel Lett² and Pierre Lefeuvre^{2,4*}

Abstract

Background Cassava (*Manihot esculenta*, Crantz), is a staple food and the main source of calories for many populations in Africa, but the plant is beset by several damaging viruses. So far, eight families of virus infecting cassava have been identified; the *Geminiviridae* (ssDNA viruses responsible for cassava mosaic disease, CMD) and *Potyviridae* (ssRNA + viruses responsible for cassava brown streak disease, CBSD) families being the most damaging to cassava in Africa. In several cassava-growing regions, the co-existence of species and strains from these two families results in a complex epidemiological situation making it difficult to correctly identify the viruses in circulation and delaying the implementation of disease management schemes. Nevertheless, the development of next generation sequencing (NGS) methods has revolutionized plant virus detection and identification. One NGS method that has been successfully used in virus detection and identification is ribodepleted RNA sequencing. Unfortunately, the relatively high cost makes it difficult to upscale this method to large epidemiological surveys and limits its adoption as a diagnostic tool.

Results Here, we develop a high-throughput sequencing protocol, named Ribo-M-Seq, that combines plant rRNA ribodepletion, cDNA synthesis, tagging with a 96 multiplexing scheme and Illumina sequencing. We evaluated the protocol on a series of cassava samples with a known assemblage of viruses. After confirming that the protocol was suitable for ribodepletion, we demonstrated it was possible to detect RNA and DNA viruses via identification of near full-size genomes. Additional phylogenetic analyses confirmed the presence of begomoviruses and ipomoviruses responsible for CMD and CBSD, respectively. We also detected a recently described ampelovirus (*Manihot esculenta*-associated virus) that was not detected in previous analyses.

Conclusions The use of the Ribo-M-Seq protocol will pave the way for large-scale sample analyses of collections with potentially complex viromes, such as those collected in the West African cassava integrated pest management program.

Keywords Cassava, Virome, Ribodepletion, RNaseH, Multiplexing, High-throughput sequencing, Virus discovery

*Correspondence:

Pierre Lefeuvre

pierre.lefeuvre@cirad.fr

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Background

Cassava (*Manihot esculenta*, Crantz) is the world's fourth-largest source of calories after rice, wheat, and maize but, most importantly, is a staple food for around 800 million people globally [1, 2]. Cassava cultivation is threatened by several diseases that cause severe yield loss [3]. In cassava-growing regions of Africa, cassava mosaic disease (CMD) and cassava brown streak disease (CBSD) are the main viral diseases causing yield loss, ranging from 40 to 100% [4, 5]. These two diseases are caused by a complex of eleven species of *Begomovirus* (ssDNA virus) [6] from the *Geminiviridae* family (*Cressdnaviricota* phylum), and two distinct species of *Ipomovirus* (ssRNA + viruses), cassava brown streak virus and Uganda cassava brown streak virus [7] from the *Potyviridae* family (*Pisuviricota* phylum), respectively. Recent studies have shown that CMD is present in all cassava-growing regions in Sub-Saharan Africa and Southern Asia. CBSD has been identified in East and Central Africa and the Comoros Archipelago [3], but is progressing towards West Africa despite control measures [8]. In addition, other viruses with a lesser or unknown impact [9, 10] have been identified, including one *Anulavirus* species (cassava Ivorian bacilliform virus) from the *Bromoviridae* family (ssRNA + virus, *Kitrinoviricota* phylum) [10] and two *Ampelovirus* species (*Manihot esculenta*-associated ampelovirus 1 and *Manihot esculenta*-associated ampelovirus 2) from the *Closteroviridae* family (ssRNA + viruses; *Kitrinoviricota* phylum) [9].

The prevention and management of plant viral diseases largely depends on the accurate identification of the viral communities responsible for the disease. However, the coexistence of several species and viral strains of these different viruses hampers the identification of circulating viruses. The absence of any canonical marker, such as the 16S gene for bacteria [11], has led virologists to develop approaches to enrich nucleic acid extracts with viral nucleic acids prior to sequencing. These next generation sequencing (NGS) methods have proved useful for the study and characterization of viromes from different sample types [12–15]. The most common approaches are virion-associated nucleic acids (VANA), double-stranded RNA (dsRNA), small interfering RNA (siRNA) and ribosomal RNA depleted total RNA [16, 17] sequencing. The latter is a credible alternative for virome characterisation and has been proved useful for the detection and discovery of RNA viruses, DNA viruses, and viroids [18, 19].

However, its use remains costly with, beside the cost of sequencing itself, costs associated to per-sample ribodepletion and sequencing library construction. Among the methods for rRNA depletion [20], RNaseH-mediated depletion (after the hybridization of reverse complementary specific DNA oligomers with rRNA, the resulting

rRNA:DNA hybrids are cleaved with RNaseH endonuclease) has been proved efficient [21]. However, this procedure is mainly implemented using high price commercial kits that limits its large-scale use in many laboratories. A second large share of the global cost of the ribosomal RNA depleted total RNA sequencing is associated with library construction, with usually one library required for one sample. Whereas methodologies exist to analyse bulk samples [22], it then requires post hoc testing to trace back identified viruses to individual samples.

The aim of this study was to implement a cost-effective high-throughput sequencing approach devised for research purpose that combine ribodepletion of total RNA extracts and molecular tagging of nucleic acids for sample multiplexing before library construction and sequencing. Here, we propose the Ribo-M-Seq protocol, a high-throughput sequencing protocol based on the ribodepletion of total RNA, cDNA synthesis and tagging of individual samples before the pooling of bulk tagged cDNAs and Illumina sequencing. We tested the effectiveness of the RNaseH enzyme for rRNA depletion and virus characterisation on cassava samples with known viral populations. We found that ribodepletion by RNaseH efficiently depleted ribosomal RNA from cassava total RNA. We were able to multiplex samples, identify DNA and RNA viruses, and obtain near-complete genomes of the target viruses. Although tested on cassava, this metagenomic protocol for virome analysis can be adapted to other plants of agronomic or historic interest whose rRNA sequences are known.

Methods

Plant samples and virus infection status

Five virus-infected dried cassava leaf samples were used as virus-infected controls (Table 1). Samples were tested for their infection status using several approaches: double-stranded RNA (dsRNA) high-throughput sequencing [9] or PCR [23] or RT-PCR [24] followed by direct Sanger sequencing of amplicons. The infection status of each sample is described in Table 1. These five samples were collected in Comoros, Madagascar, Mayotte and Reunion between 2011 and 2016. Cassava leaves from uninfected vitroplants, frozen at -80°C , were used as negative control.

Molecular analysis of the cassava viromes

Total RNA was extracted using the RNeasy Plus Kit (Qiagen, Les Ulis, France) according to the manufacturer's instructions. Total RNA quantity was assessed with the Qubit fluorometer (Thermo Fisher Scientific Inc., Waltham, MA) using the RNA HS Assay kit (Thermo Fisher Scientific, Illkirch, France).

Table 1 List of cassava samples used in the study with details on previous virus detections

Samples	Origin/Site	Collection date	Virus name (Acronyme)	Taxonomic name	Genome	Family/Genus	Virus detection	References
Healthy vitro-plant	Reunion Island/ Saint-Pierre	20/01/2022	–	–	–	–	–	–
293MG040711	Madagascar/ Diana	04/07/2011	African cassava mosaic virus (ACMV)	<i>Begomovirus manihotis</i>	ssDNA	<i>Geminiviridae/ Begomovirus</i>	PCR / Direct Sanger Sequencing	Harimalala et al., 2015
			East African cassava mosaic Cameroon virus (EACMCV)	<i>Begomovirus manihotis-cameroonense</i>			PCR / Direct Sanger Sequencing	
			East African cassava mosaic Kenya virus (EACMKV)	<i>Begomovirus manihotiskeny-aense</i>			PCR / Direct Sanger Sequencing	
HAY1.3	Comoros/ Havrara	15/09/2016	Cassava brown streak virus (CBSV)	<i>Ipomovirus bru-nusmanihotis</i>	ssRNA	<i>Potyviridae/ Ipomovirus</i>	RT-PCR / Direct Sanger Sequencing	Azali et al., 2017
6 mois Blanc	Mayotte/Dem-beni	26/03/2015	East African cassava mosaic virus (EACMV)	<i>Begomovirus manihotisafri-caense</i>	ssDNA	<i>Geminiviridae/ Begomovirus</i>	dsRNAs	Kwibuka et al., 2021
			Manihot esculenta-associated virus 1 (MEaV-1)	na	ssRNA	<i>Closteroviridae/ Ampelovirus</i>		
			Manihot esculenta-associated virus 2 (MEaV-2)	na	ssRNA	<i>Closteroviridae/ Ampelovirus</i>		
CRE11	Reunion Island/ Bassin Plat	25/02/2015	Manihot esculenta-associated virus 1 (MEaV-1)	na	ssRNA	<i>Closteroviridae/ Ampelovirus</i>	dsRNAs	Kwibuka et al., 2021
HEL 3.1	Comoros/ Helindje	15/09/2016	Ugandan cassava brown streak virus (UCBSV)	<i>Ipomovirus manihotis</i>	ssRNA	<i>Potyviridae/ Ipomovirus</i>	RT-PCR / Direct Sanger Sequencing	Azali et al., 2017

A protocol for high-throughput sequencing based on ribodepletion of total RNA, dsDNA synthesis and tagging was implemented for cassava virome analysis (Fig. 1). Ribodepletion was achieved via cleavage of rRNA hybridised with specific DNA probes using RNaseH [25]. A total of 273 DNA oligomers were designed on the basis of rRNA cassava sequences of reference cassava genome v8.1 (GCF_001659605.2). The oligomers were designed as described by Phelps et al. [25] using the Oligo-ASST Web tool (<https://mtleelab.pitt.edu/oligo>), resulting in a pool of 273 unique oligomers. Ribodepletion by RNaseH was performed as described by Phelps et al. [25] with slight modifications: the total amount of RNA per sample was reduced to 100 ng and the final concentration of oligomers was 0.036 μ M. The RNA–DNA hybrids were digested using 10 U of thermostable RNase H (EURx, Gdańsk, Poland) at 65 °C for 10 min in a 20 μ L volume. After digestion, the sample was purified using Mag-Bind total pure next-generation sequencing (NGS) beads (1.8X,

Omega Bio-Tek, Tebubio, Le Perray en Yvelines, France) and ribodepleted RNAs were eluted in 35 μ L of nuclease-free water. Two control treatments were used: the first consisted of total RNA direct use without any ribodepletion treatment and the second consisted of total RNA treated using RNaseH but in the absence of of rRNA specific complementary oligomers. Whereas the first control treatment was applied to every samples, this second control treatment was applied to the healthy cassava control sample and the 6 mois Blanc sample (Table 1). A total of 14 sample-treatment combinations was analysed.

Purified ribodepleted RNA was used for complementary DNA (cDNA) synthesis and tagging as described by François et al. [26], except purification which was done with Mag-Bind total pure next-generation sequencing (NGS) beads (1.8X, Omega Bio-Tek, Tebubio, Le Perray en Yvelines, France). Using that protocol, DNA amplicon sets with unique tags of 24 nt on both extremities are obtained (see François et al. [26] for details on tag

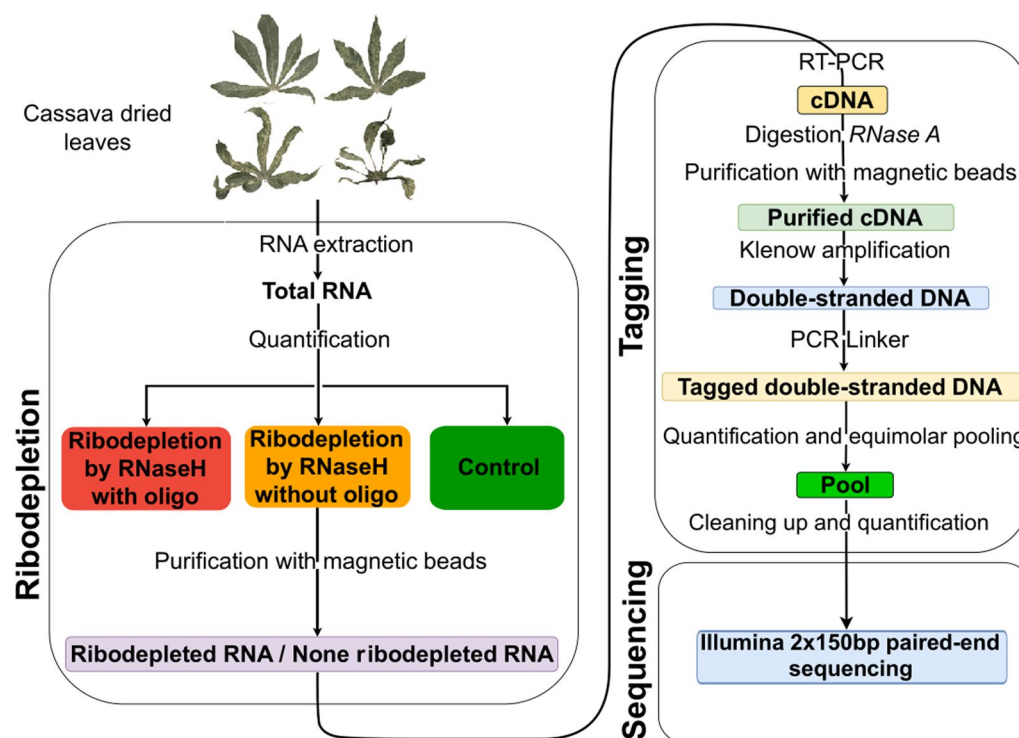


Fig. 1 Schematic representation of the Ribodepletion-Multiplexing-Sequencing protocol with the ribodepletion, tagging and sequencing steps

sequences). Each sample was treated in triplicate with three different tags from the 96. Amplicons obtained were quantified using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, Illkirch, France) before equimolar pooling. The amplicon pool was then cleaned up using Mag-Bind total pure next-generation sequencing (NGS) beads (0.65X), and quantified using Qubit dsDNA HS Assay Kit. The pool was sent for 2×150 bp paired-end sequencing on an Illumina NovaSeq 6000 sequencer at Eurofins Genomics (Ebersberg, Germany). Amplicon pool was checked using High Sensitivity D5000 ScreenTape for Agilent Tapestation (Additional Fig. 1). Illumina sequencing library was constructed by the manufacturer with their PCR-based protocol. A 10% PhiX spike-in was used during sequencing.

Bioinformatics analysis

After Illumina sequencing, reads were demultiplexed and the 24 nt tags were removed using Cutadapt v3.5 [27]. The double indexed reads were quality controlled using Trimmomatic v0.35 [28], over a sliding window of five bases with an average quality of 20. Adapters were removed, and poor quality and/or short reads (fewer than 100 bases) were discarded. The cleaned reads were then used for similarity searches against a database of virus sequences from NCBI RefSeq (obtained in October 2022, release 213) and the cassava reference genome with

MMseqs2 [29]. The total number of reads assigned to the cassava genome, rRNA, and viruses were recorded. On a per sample basis, reads were de novo assembled using SPAdes v3.13.0 [30] and mapped back against the assembled contigs using bwa-mem2 v2.2.1 [31]. Mapping statistics were determined using SAMtools v1.18 [32]. The contigs and unmapped reads were then used in similarity searches against the above mentioned database using MMseqs2. For sequences identified as viruses, a second similarity search analysis was performed using BLASTn and BLASTx against the RefSeq viral database using an E-value of 10^{-4} as the cut-off threshold value for significant hits.

Viral contigs of more than 500 nucleotides (nt) were sorted by virus family before being aligned using MAFFT v7.453 [33] against representative genomes of this family obtained from GenBank in August 2023. Maximum-likelihood phylogenetic trees were inferred with FastTree v2.3 [34] using the general time reversible and gamma parameters. Branch supports were tested using the Shimodaira–Hasegawa procedure. Phylogenetic trees were edited using the ape R package [35].

In order to estimate the coverage of the largest viral contigs in relation to the number of sequenced reads per sample, sub-samplings of the contigs coverage data were performed. To this end, the actual number of reads mapped per position of contigs representing full size or

nearly full size of viral genomes were sub-sampled 100 times for sets of decreasing sequencing efforts. Sequencing depth (the number of times a position was covered with a read) and breadth of the coverage (the proportion of the genome covered with a read) were calculated for each subsample.

Results and discussion

Effectiveness of ribodepletion

After demultiplexing raw reads, it was apparent that a large fraction of the reads (47%) presented with mismatching tags or had at least one of the two reads without identifiable tag (5%). Probable high index-switching rates associated to the use of PCR for sequencing library construction might be at the root of such issue. Comparable results were reported in other studies using similar library construction and sequencing procedure [36–38]. The index switching are known to results from the formation of chimera during bulk amplification of tagged amplicons during library index PCR [36]. The use of PCR during library preparation from amplicons should thus be avoided for better results. In order to lie on the side of conservatism, we choose to only consider for further analysis the fraction of reads pairs that presented with matching tags (48% of the raw reads). After quality control of pairs with matching tags, 75% of the demultiplexed reads remained and the final number of reads associated to each of the studied sample/treatment combination varied from 3.9 to 21.9 million with a mean of 10.3 million.

In order to evaluate the effectiveness of the ribodepletion, the clean reads were used for an initial global classification (Fig. 2). Reads were classified as either cassava rRNA sequences, cassava genomic sequences or virus sequences. For the healthy vitroplant control without ribodepletion, the percentage of reads associated with rRNAs and cassava genome were 95.7% and 4.8%, respectively. These proportions were largely similar (82.0% and 2.6% for rRNA and genome, respectively) for the second control with RNaseH treatment but without probes. Conversely, after RNaseH treatment, the percentage of reads from the rRNA and cassava genome were 0.5% and 95.6% respectively, indicative of a near-complete rRNA depletion. Similar trends were obtained for the other samples with a large decrease of rRNA reads after RNaseH treatments in comparison to the control without treatment or RNaseH treatment without probes. While proportions of cassava genome sequences and rRNA ranged from 1.6 to 9.1% and 63.0 to 97.1% respectively for controls, no samples gave more than 8.8% of rRNA reads after RNaseH treatment. However, the proportion of reads attributed to the cassava genome increased after ribodepletion, ranging from 6.0 to 29.8% (average 16.7%). The remaining

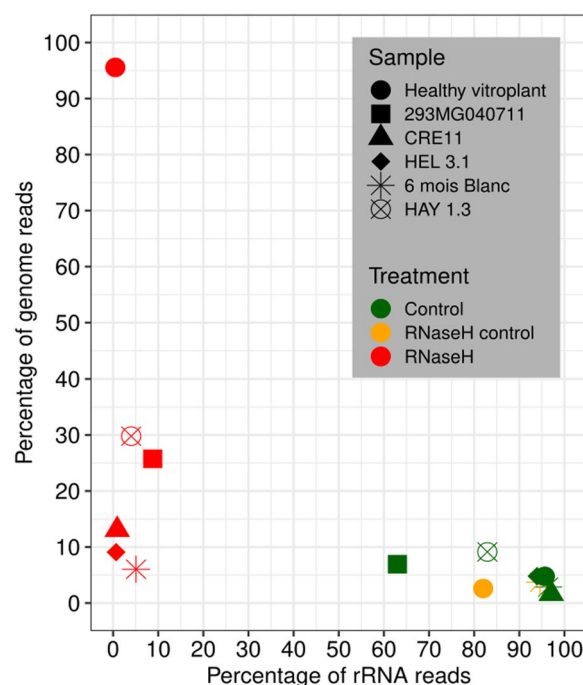


Fig. 2 Percentage of reads assigned to cassava ribosomal RNAs (x-axis) and other cassava genome reads (y-axis) for each sample under the different treatments applied, as per key at the top right of the figure

sequences were unclassified (65.4% to 90.2%). It must be noted that such a large proportion of unclassified sequences was not observed for the healthy cassava control (mean: 93.7% of classified reads). Further attempts to classify these reads revealed hits with significant proportions for fungal RNAs and rRNAs (data not shown). Whereas the ribodepletion protocol presented here ensures efficient plant rRNA removal from total RNA as showed in previous studies [21, 25], our results also highlighted the importance of sample conservation and the limitations of using relatively old samples. Although we were able to extract and sequence RNA from dehydrated samples conserved at room temperature for up to eleven years, a large fraction of fungal RNA was obtained from the samples despite the absence of visible fungi growth.

Estimation of the background

Estimating the proportion of viral reads from the negative control requires estimating the mean background contamination [39]. Analysis of the ~7.6 M reads obtained after quality control of the negative control allowed us to assign 30 reads to viral genomic sequences, with a maximum of 20 reads to members of the *Potyviridae* family (Table 2). This represented less than four viral reads per million sequenced reads (less than three for members of *Potyviridae*). Note that establishing an exact threshold to

Table 2 Classification of viral sequence at the family level for each cassava sample after ribodepletion

Sample	Total number of reads	<i>Closteroviridae</i> members*	<i>Geminiviridae</i> members*	<i>Potyviridae</i> members*	Total number of viral reads*
Healthy vitroplant	7,637,666	1 (0)	9 (1)	20 (3)	30 (4)
293MG040711	9,693,254	460 (47)	2311 (239)	0 (0)	2771 (286)
HAY 1.3	4,530,258	492 (109)	111 (25)	4673 (1032)	5276 (1166)
6 mois Blanc	7,477,360	22 (3)	290 (39)	5 (1)	317 (43)
CRE11	11,705,482	48 (4)	93 (8)	1 (0)	142 (12)
HEL 3.1	4,456,320	0 (0)	12 (3)	8 (2)	20 (5)

* Rounded number of reads per million sequenced reads are indicated in parenthesis

determine positivity is not an easy feat using NGS data and more controls are required for a thorough statistical estimation of this threshold [39]. A negative control made of healthy cassava herbarium in addition to the fresh cassava control would certainly have proved informative. However, based on the above estimation of the number of viral reads detected from the negative control, a conservative value of 100 reads per million sequenced reads (1 in 10,000 or 0.01%) would be used to filter our results, a threshold in line with reports from positive samples analysed using a similar approach [16, 40–42].

Taxonomic assignments and characterisation of plant viruses

Congruent with our background knowledge of the viruses infecting the tested samples (Table 1), reads were mainly assigned to viruses of the *Closteroviridae* (ssRNA+), *Geminiviridae* (circular ssDNA), and *Potyviridae* (ssRNA+) families (Table 2). For sample 293MG040711 infected by three begomoviruses (African cassava mosaic virus, ACMV; East African cassava mosaic Cameroon virus, EACMCV and East African cassava mosaic Kenya virus, EACMKV), the presence of the begomoviruses previously characterised using the RCA-RFLP method [23] was confirmed. A total of 2,311 begomovirus reads were detected (513 ACMV reads, 1,529 EACMCV reads and 233 EACMKV reads). In addition to virus detection, we also obtained contigs of ACMV (176 to 1,146 nt), EACMCV (173 to 2,698 nt) and EACMKV (456 to 1,244 nt). Five contigs of more than 500 nt were used for phylogenetic inference. These contigs were clustered (with nucleotide identities ranging from 94.2 to 100%) with sequences of other isolates obtained from Madagascar (Additional Fig. 2). Unexpectedly, 443 reads of *Manihot esculenta* associated ampelovirus 2 (genus *Ampelovirus*, assembled in ten contigs of 238 to 1,830 nt) were also obtained from the sample. The contigs clustered with isolates of *Manihot esculenta*-associated virus

(Additional Fig. 3), which was also identified in Madagascar. It is important to notice that previous analyses of the sample focused on CMGs and no ampelovirus indexing was thus carried out. Besides highlighting the diversity and distribution of the cassava ampeloviruses, this also demonstrates that the NGS protocol used is suitable for the co-detection of RNA and DNA viruses.

For the HAY1.3 sample, the CBSV (genus *Ipomovirus*) was previously detected by RT-PCR (Table 1). This detection was confirmed in our analysis, with a total of 4,673 reads assigned to this species. These reads were assembled into three CBSV contigs including one of 8,582 nt, almost the entire length of the closest isolate whose full genome is available (MK103393; 9,002 nt). The phylogeny of the CBSV (Additional Fig. 4) revealed that the contig was closely related (maximum nucleotide identity 95.8%) to three other isolates obtained from samples collected in Grande Comore [24]. Finally, as for sample 293MG040711, unexpected ampelovirus reads were obtained from sample HAY1.3 (N=459) and six contigs of more than 500 nt were assembled. The associated phylogeny shows that these contigs were most closely related to other isolates of *Manihot esculenta*-associated ampelovirus 1 from Madagascar and Mayotte [9]. The details of the contigs are presented in Additional Table 1.

The importance of sample preservation for virus detection

For 6 mois Blanc sample, from which sequences of ampeloviruses and begomoviruses had previously been obtained, we could only confidently confirm the detection of begomoviruses with 272 reads. However, no medium size contigs could be assembled and no further classification were attempted. The last two samples, CRE11 and HEL3.1, while giving some virus reads, had counts of similar magnitude as the healthy control and as such were not considered for further analysis. We were thus unable to confirm the previous viral identification for these three samples. The fact that these three samples

had the lowest proportion of classified reads (maximum of 14% in comparison to ~34% for both 293MG040711 and HAY1.3) points again to the importance of sample preservation for accurate analysis, most importantly when dealing with low titer viruses that may be difficult to detect [43, 44]. Our samples were collected between 2011 and 2016 and were preserved in envelopes in a herbarium. High susceptibility of RNA to hydrolytic attack [45] and long-term storage of dried leaves, known to be associated to damage of nucleic acids [46], might have had a negative impact on virus identification [47]. Comprehensive RNA quality control would thus be recommended before using the described protocol.

Influence of sequencing depth on viral genome coverage

In order to evaluate the sequencing effort required for virus characterisation, we choose to thoroughly sequence each sample to later estimate the actual number of samples that could be multiplexed while maintaining the ability to identify the viruses in these samples. For samples 293MG040711 and HAY1.3, the breadth of coverage (i.e. the proportion of the viral genome that is covered with reads) was calculated at a sequencing depth of 10X (i.e. meaning that a given position has to be covered with at least ten reads to be considered) for sets of subsampled reads. We obtained the distribution of coverage percentage of the genome for each species of virus depending on the number of sequenced bases (Fig. 3). For sample 293MG040711, the breadth of coverage of CMGs DNA-A and DNA-B components were both above 90% and for the ampelovirus genome this figure was 88% (Fig. 3A). For sample HAY 1.3, the breadth of coverage was 46%, 37% and 28% for CMGs DNA-A, CMGs DNA-B and ampelovirus genomes, respectively. It was 84% for the CBSV genome (Fig. 3B). Not all the viruses benefited from the same efficiency of characterisation; these differences could be attributed to variations in abundance [48, 49] and/or variations in RNA stability [50]. As we were not able to obtain full genome 10X coverage for any of the analysed viruses, the significance of the results remain limited. However, for CMGs DNA-A and DNA-B sequences from 293MG040711, the curves tended to plateau, indicating that 100% breadth of coverage may not be achievable for these viruses. Conversely, steady increases in breadth of coverage were observed for the ampelovirus genome from 293MG040711 and for all viruses identified from the HAY1.3 sample. This latter observation indicates that the addition of new reads would improve virome characterisation. As such, any increases in the number of multiplexed samples, thus

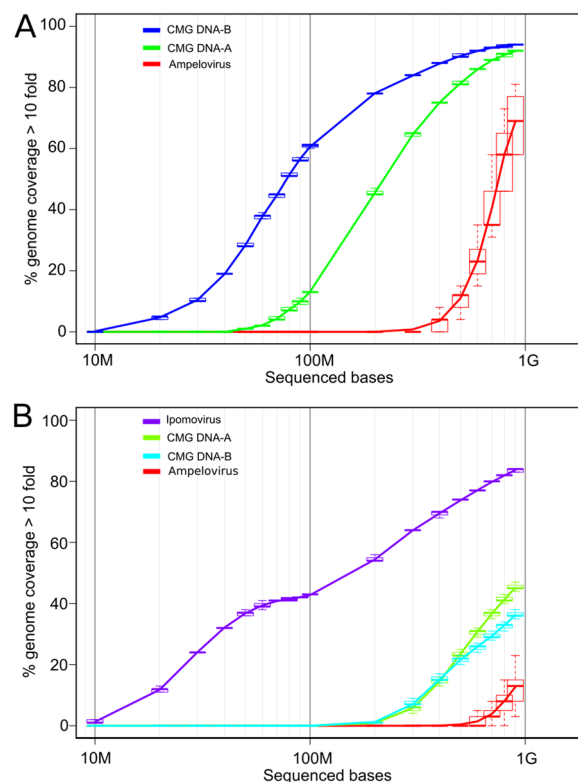


Fig. 3 Intraplotation of the breadth of coverage at a 10X coverage (y-axis) for representative genomes of ampeloviruses, cassava geminiviruses DNA-A and DNA-B components and ipomoviruses according to the number of sequenced bases (x-axis, in Log10 scale) for sample 293MG040711 (A) and HAY 1.3 (B)

reducing the per-sample read numbers, would decrease our ability to characterize viral genomes. The multiplexing/coverage trade-off is delicate and depends on the scientific goal of the experiment. For virus detection, without any a priori, the sequencing effort in this study was sufficient to improve on previous knowledge of the virome of some samples. However, for poorer quality samples, analysis was unsuccessful. The poor quality of the samples that we analysed limited the sequencing quality, resulting in, at best, only a third of the sequences being successfully catalogued. Given that for the healthy cassava control, obtained from fresh material, 84% of the total reads were classified, a three-fold increase in usable reads would be expected in virome characterisation, if fresh samples were used. This would convert to ~42 samples analysed in a run (14 combinations of samples and treatments were analysed here) that could conveniently be limited to 32 to treat samples in triplicates and employ a 96-tag scheme.

Conclusion

The originality of the procedure lies in the combination of two widely used protocols for ribodepletion and amplicon tagging in order to make virus detection from total RNA extracts more affordable. Whereas our work demonstrates that ribodepletion with RNaseH effectively removed most rRNA from total cassava RNA, our results also point to the importance of sample conservation for effective ribodepletion and virus detection. The strategy made it possible to detect RNA and DNA viruses and obtain contigs with near full-length viral genomes of target viruses. Although specific probe design has to be conducted depending on the plant species analysed, the procedure remains an inexpensive alternative that can be adapted to any plant whose rRNA sequences are known. With a per-sample ribodepletion and tagging price of around 18€, cost savings are achievable on both ribodepletion and multiplexing. The ability to multiplex up to 32 samples in a single library before sequencing in a single lane makes this an attractive alternative method of virus detection and characterisation for research studies in plant virus epidemiology.

Abbreviations

ACMBFV	African cassava mosaic Burkina Faso virus
ACMV	African cassava mosaic virus
CBSD	Cassava Brown Streak Disease
CBSV	Cassava brown streak virus
CMD	Cassava Mosaic Disease
CMGs	Cassava mosaic Geminiviruses
CMMGV	Cassava mosaic Madagascar virus
dsRNA	Double-stranded RNA
EACMCV	East African cassava mosaic Cameroon virus
EACMKV	East African cassava mosaic Kenya virus
EACMMV	East African cassava mosaic Malawi virus
EACMV	East African cassava mosaic virus
EACMZV	East African cassava mosaic Zanzibar virus
GLRaV1	Grapevine leafroll-associated virus 1
GLRaV13	Grapevine leafroll-associated virus 13
GLRaV3	Grapevine leafroll-associated virus 3
GLRaV4	Grapevine leafroll-associated virus 4
ICMV	Indian cassava mosaic virus
LChV2	Little cherry virus 2
MEaV	Manihot esculenta-associated ampelovirus
NGS	Next Generation Sequencing
PAVA	Pistachio ampelovirus A
PBNPaV	Plum bark necrosis stem pitting-associated virus
PMWaV1	Pineapple mealybug wilt-associated virus 1
PMWaV2	Pineapple mealybug wilt-associated virus 2
PMWaV3	Pineapple mealybug wilt-associated virus 3
RCA-RFLP	Rolling Circle Amplification-restriction fragment length polymorphism
Ribo-M-Seq	Ribodepletion-Multiplexing-Sequencing
rRNA	Ribosomal RNA
SACMV	South African cassava mosaic virus
siRNA	Small interfering RNA
SLCMV	Sri Lankan cassava mosaic virus
circular ssDNA	Circular single-stranded DNA
ssDNA	Single-stranded DNA

ssRNA	Single-stranded RNA
UCBSV	Uganda cassava brown streak virus
VANA	Virion Associated Nucleic Acid
YaV1	Yam asymptomatic virus 1

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12985-025-02634-9>.

Additional file 1.
Additional file 2.
Additional file 3.
Additional file 4.
Additional file 5.

Acknowledgements

The authors would like to thank Camille Gendron for providing samples of healthy cassava vitroplants.

Author contributions

D.H.O.: Conceptualization, Formal analysis, Investigation, Data Curation, Writing—Original Draft, Writing—Review & Editing. J.S.P.: Conceptualization, Writing—Review & Editing, Funding acquisition. M.H.: Conceptualization, Investigation. F.T.: Conceptualization, Writing—Review & Editing, Project administration. J.M.L.: Conceptualization, Validation, Resources, Writing—Review & Editing, Project administration, Supervision. P.L.: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing—Review & Editing, Supervision. All authors reviewed and approved the final manuscript.

Funding

This research was funded by the Bill and Melinda Gates Foundation and the United Kingdom Foreign, Commonwealth, and Development Office (FCDO; INV-002969; grant no. OPP1212988) to the Central and West African Virus Epidemiology (WAVE) Program for root and tuber crops, Université Félix Houphouët-Boigny (UFHB), the European Regional Development Fund (FEDER), the Région Réunion and CIRAD.

Availability of data and materials

Sequence data used and analysed during the current study are available at the NCBI Short read archive under the BioProject PRJNA1174894.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹The Central and West African Virus Epidemiology (WAVE) for Food Security Program, Pôle Scientifique Et d'Innovation, Université Félix Houphouët-Boigny (UFHB), Abidjan 22 BP 582, Côte d'Ivoire. ²CIRAD, UMR PVBMT F-97410, St Pierre, La Réunion, France. ³UFR Biosciences, Université Félix Houphouët-Boigny (UFHB), Abidjan 22 BP 582, Côte d'Ivoire. ⁴CIRAD, UMR PVBMT, Department of Plant Protection, College of Agriculture, Can Tho University, Can Tho City, Vietnam.

Received: 14 November 2024 Accepted: 16 January 2025
Published online: 05 February 2025

References

- Landicho D, Balendres MA. Possible incursion of cassava virus diseases: risks and potential threats to the Philippine cassava industry. *Arch Phytopathol Plant Prot.* 2022;55:1725–49. <https://doi.org/10.1080/03235408.2022.2110662>.
- Otun S, Escrib A, Achilonu I, Rauwane M, Lerma-Escalera JA, Rubén Morones-Ramírez J, et al. The future of cassava in the era of biotechnology in Southern Africa. *Crit Rev Biotechnol.* 2023;43:594–612. <https://doi.org/10.1080/07388551.2022.2048791>.
- Robson F, Hird DL, Boa E. Cassava brown streak: a deadly virus on the move. *Plant Pathol.* 2023;73:221–41. <https://doi.org/10.1111/ppa.13807>.
- Bisimwa E, Walangululu J, Bragard C. Cassava mosaic disease yield loss assessment under various altitude agroecosystems in the sudKivu region. *Democr Repub Congo Trop.* 2015;33:101–10.
- Kwibuka Y, Nyirakanani C, Bizimana JP, Bisimwa E, Brostaux Y, Lassois L, et al. Risk factors associated with cassava brown streak disease dissemination through seed pathways in Eastern D.R. Congo *Front Plant Sci.* 2022;13:1–18. <https://doi.org/10.3389/fpls.2022.803980>.
- Crespo-Bellido A, Hoyer JS, Dubey D, Jeannot RB, Duffy S. Interspecies recombination has driven the macroevolution of Cassava Mosaic Begomoviruses. *J Virol.* 2021;95(17):10–1128. <https://doi.org/10.1128/jvi.00541-21>.
- Mbewe W, Mukasa S, Ochwo-Ssemakula M, Sseruwagi P, Tairo F, Ndunguru J, et al. Cassava brown streak virus evolves with a nucleotide substitution rate that is typical for the family Potyviridae. *Virus Res.* 2024;346: 199397. <https://doi.org/10.1016/j.virusres.2024.199397>.
- Rey C, Vanderschuren H. Cassava mosaic and brown streak diseases: current perspectives and beyond. *Annu Rev Virol.* 2017;4:429–52. <https://doi.org/10.1146/annurev-virology-101416-041913>.
- Kwibuka Y, Bisimwa E, Blouin AG, Bragard C, Candresse T, Faure C, et al. Novel ampeloviruses infecting cassava in central africa and the south-west indian ocean islands. *Viruses.* 2021;13:1–17. <https://doi.org/10.3390/v13061030>.
- Scott SW, MacFarlane SA, McGavin WJ, Fargette D. Cassava ivoirien bacilliform virus is a member of the genus anulavirus. *Arch Virol.* 2014;159:2791–3. <https://doi.org/10.1007/s00705-014-2086-3>.
- Srinivasan R, Karaoz U, Volegova M, MacKichan J, Kato-Maeda M, Miller S, et al. Use of 16S rRNA gene for identification of a broad range of clinically relevant bacterial pathogens. *PLoS ONE.* 2015;10:1–22. <https://doi.org/10.1371/journal.pone.0117617>.
- Bejerman N, Roumagnac P, Nemchinov LG. High-throughput sequencing for deciphering the virome of alfalfa (*Medicago sativa* L.). *Front Microbiol.* 2020;11: 553109. <https://doi.org/10.3389/fmicb.2020.553109>.
- Mutuku JM, Wamoni FO, Mukeshimana G, Njuguna J, Wamalwa M, Choi S-K, et al. Metagenomic analysis of plant virus occurrence in common bean (*Phaseolus vulgaris*) in central Kenya. *Front Microbiol.* 2018;9:2939. <https://doi.org/10.3389/fmicb.2018.02939>.
- Schönegger D, Moubset O, Margaria P, Menzel W, Winter S, Roumagnac P, et al. Benchmarking of virome metagenomic analysis approaches using a large, 60+ members, viral synthetic community. *J Virol.* 2023;97(11):e01300–e1323. <https://doi.org/10.1128/jvi.01300-23>.
- Wainaina JM, Ateka E, Makori T, Kehoe MA, Boykin LM. A metagenomic study of DNA viruses from samples of local varieties of common bean in Kenya. *PeerJ.* 2019;7: e6465. <https://doi.org/10.7717/peerj.6465>.
- Gaafar YZA, Ziebell H. Comparative study on three viral enrichment approaches based on RNA extraction for plant virus/viroid detection using high-throughput sequencing. *PLoS ONE.* 2020;15: e0237951. <https://doi.org/10.1371/journal.pone.0237951>.
- Roossinck MJ, Martin DP, Roumagnac P. Plant virus metagenomics: Advances in virus discovery. *Phytopathology.* 2015;105:716–27. <https://doi.org/10.1094/PHYTO-12-14-0356-RVW>.
- Cobbin JC, Charon J, Harvey E, Holmes EC, Mahar JE. Current challenges to virus discovery by meta-transcriptomics. *Curr Opin Virol.* 2021;51:48–55. <https://doi.org/10.1016/j.coviro.2021.09.007>.
- Haegeman A, Foucart Y, De Jonghe K, Goedefroid T, Al Rwahnih M, Boonham N, et al. Looking beyond virus detection in RNA sequencing data: lessons learned from a community-based effort to detect cellular plant pathogens and pests. *Plants.* 2023;12(1):2139. <https://doi.org/10.3390/plants12112139>.
- Adiconis X, Borges-Rivera D, Satija R, Deluca DS, Busby MA, Berlin AM, et al. Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nat Methods.* 2013;10:623–9. <https://doi.org/10.1038/nmeth.2483>.
- Baldwin A, Morris AR, Mukherjee N. An easy, cost-effective, and scalable method to deplete human ribosomal RNA for RNA-seq. *Curr Protoc.* 2021;1:1–13. <https://doi.org/10.1002/cpz1.176>.
- Fowkes AR, McGreig S, Pufal H, Duffy S, Howard B, Adams IP, et al. Integrating high throughput sequencing into survey design reveals turnip yellows virus and soybean dwarf virus in pea (*Pisum sativum*) in the united kingdom. *Viruses.* 2021;13:2530. <https://doi.org/10.3390/v13122530>.
- Harimalala M, Chiroleu F, Giraud-carrier C, Hoareau M, Zinga I, Randria-mampianina J, et al. Molecular epidemiology of cassava mosaic disease in Madagascar. *Plant Pathol.* 2015;64(3):501–7. <https://doi.org/10.1111/ppa.12277>.
- Azali HA, Maillot V, Cassam N, Chesneau T, Soulezelle J, Scussel S, et al. Occurrence of cassava brown streak disease and associated Cassava brown streak virus and Ugandan cassava brown streak virus in the Comoros Islands. *New Dis Reports.* 2017;36(1):19–19. <https://doi.org/10.5197/j.2044-0588.2017.036.019>.
- Phelps WA, Carlson AE, Lee MT. Optimized design of antisense oligomers for targeted rRNA depletion. *Nucleic Acids Res.* 2021;49(1):1–12. <https://doi.org/10.1093/nar/gkaa1072>.
- François S, Filloux D, Fernandez E, Ogliastro M, Roumagnac P. 2018 Viral Metagenomics Approaches for High-Resolution Screening of Multiplexed Arthropod and Plant Viral Communities. In: Pantaleo V, Chiumenti M, (Eds.) *Viral Metagenomics Methods Protoc.* Springer. Newyork. https://doi.org/10.1007/978-1-4939-7683-6_7.
- Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 2011;17:10–2. <https://doi.org/10.14806/ej.17.1.200>.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114–20. <https://doi.org/10.1093/bioinformatics/btu170>.
- Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol.* 2017;35:1026–8. <https://doi.org/10.1038/nbt.3988>.
- Prijbelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes de novo assembler. *Curr Protoc Bioinforma.* 2020;70: e102. <https://doi.org/10.1002/cpbi.102>.
- Vasimuddin M, Misra S, Li H, Aluru S. 2019 Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. *IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, Rio de Janeiro, Brazil, 314–324. <https://doi.org/10.1109/IPDPS.2019.00041>.
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *Gigascience.* 2021;10(2):giab008. <https://doi.org/10.1093/gigascience/giab008>.
- Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30:772–80. <https://doi.org/10.1093/molbev/mst010>.
- Price MN, Dehal PS, Arkin AP. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE.* 2010;5(3): e9490. <https://doi.org/10.1371/journal.pone.0009490>.
- Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics.* 2019;35:526–8. <https://doi.org/10.1093/bioinformatics/bty633>.
- Bohmann K, Elbrecht V, Carøe C, Bista I, Leese F, Bunce M, et al. Strategies for sample labelling and library preparation in DNA metabarcoding studies. *Mol Ecol Resour.* 2022;22:1231–46. <https://doi.org/10.1111/1755-0998.13512>.
- Esling P, Lejzerowicz F, Pawlowski J. Accurate multiplexing and filtering for high-throughput amplicon-sequencing. *Nucleic Acids Res.* 2015;43:2513–24. <https://doi.org/10.3390/v13122530>.
- Carøe C, Bohmann K. Tagsteady: a metabarcoding library preparation protocol to avoid false assignment of sequences to samples. *Mol Ecol Resour.* 2020;20:1620–31. <https://doi.org/10.1111/1755-0998.13227>.

39. Massart S, Adams I, Al RM, Baeyen S, Bilodeau J, Blouin AG, et al. Guidelines for the reliable use of high throughput sequencing technologies to detect plant pathogens and pests. *Peer Community J*. 2022;2:62. <https://doi.org/10.24072/pcjournal.181>.
40. Pecman A, Kutnjak D, Gutiérrez-Aguirre I, Adams I, Fox A, Boonham N, et al. Next generation sequencing for detection and discovery of plant viruses and viroids: comparison of two approaches. *Front Microbiol*. 2017;8:1–10. <https://doi.org/10.3389/fmicb.2017.01998>.
41. Pecman A, Adams I, Gutiérrez-Aguirre I, Fox A, Boonham N, Ravnikar M, et al. Systematic comparison of nanopore and illumina sequencing for the detection of plant viruses and viroids using total RNA sequencing approach. *Front Microbiol*. 2022;13:1–14. <https://doi.org/10.3389/fmicb.2022.883921>.
42. Malapi-Wight M, Adhikari B, Zhou J, Hendrickson L, Maroon-Lango CJ, McFarland C, et al. Hts-based diagnostics of sugarcane viruses: seasonal variation and its implications for accurate detection. *Viruses*. 2021;13(8):1627. <https://doi.org/10.3390/v13081627.47>.
43. Maclot F, Candresse T, Filloux D, Malmstrom CM, Roumagnac P, van der Vlugt R, et al. Illuminating an ecological blackbox: using high throughput sequencing to characterize the plant virome across scales. *Front Microbiol*. 2020;11: 578064. <https://doi.org/10.3389/fmicb.2020.578064>.
44. Gallo Y, Marín M, Gutiérrez P. Detection of RNA viruses in Solanum quitoense by high-throughput sequencing (HTS) using total and double stranded RNA inputs. *Physiol Mol Plant Pathol*. 2021;113: 101570. <https://doi.org/10.1016/j.pmpp.2020.101570>.
45. Campbell MK, Farrell SHO, McDougal OM. *Biochemistry*. 9th ed. Boston, USA: Cengage Learning; 2018.
46. Staats M, Cuenca A, Richardson JE, van Ginkel RV, Petersen G, Seberg O, et al. DNA damage in plant herbarium tissue. *PLoS ONE*. 2011;6(12): e28448. <https://doi.org/10.1371/journal.pone.0028448>.
47. Mark D, Tairo F, Ndunguru J, Kweka E, Saggaf M, Bachwenkizi H, et al. Assessing the effect of sample storage time on viral detection using a rapid and cost-effective CTAB-based extraction method. *Plant Methods*. 2024;20:1–16. <https://doi.org/10.1186/s13007-024-01175-6>.
48. Charlebois RL, Sathiamoorthy S, Logvinoff C, Gisonni-Lex L, Mallet L. Ng SHS. sensitivity and breadth of detection of high-throughput sequencing for adventitious virus detection. *npj Vaccines*. 2020;5:1–8. <https://doi.org/10.1038/s41541-020-0207-4>.
49. Ogunbayo AE, Sabiu S, Nyaga MM. Evaluation of extraction and enrichment methods for recovery of respiratory RNA viruses in a metagenomics approach. *J Virol Methods*. 2023;314: 114677. <https://doi.org/10.1016/j.jviromet.2023.114677>.
50. Zhang K, Hodge J, Chatterjee A, Moon TS, Parker KM. Duplex structure of double-stranded RNA provides stability against hydrolysis relative to single-stranded RNA. *Environ Sci Technol*. 2021;55:8045–53. <https://doi.org/10.1021/acs.est.1c01255>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.