



HAL
open science

Nanopore Sequencing Is a Credible Alternative to Recover Complete Genomes of Geminiviruses

Selim Ben Chehida, Denis Filloux, Emmanuel Fernandez, Oumaima Moubset, Murielle Hoareau, Charlotte Julian, Laurence Blondin, Jean-Michel Lett, Philippe Roumagnac, Pierre Lefevre

► **To cite this version:**

Selim Ben Chehida, Denis Filloux, Emmanuel Fernandez, Oumaima Moubset, Murielle Hoareau, et al.. Nanopore Sequencing Is a Credible Alternative to Recover Complete Genomes of Geminiviruses. *Microorganisms*, 2021, 9 (5), pp.903. 10.3390/microorganisms9050903 . hal-03276551

HAL Id: hal-03276551

<https://hal.inrae.fr/hal-03276551>

Submitted on 2 Jul 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Article

Nanopore Sequencing Is a Credible Alternative to Recover Complete Genomes of Geminiviruses

Selim Ben Chehida ¹, Denis Filloux ^{2,3}, Emmanuel Fernandez ^{2,3}, Oumaima Moubset ^{2,3}, Murielle Hoareau ¹, Charlotte Julian ^{2,3}, Laurence Blondin ^{2,3}, Jean-Michel Lett ¹, Philippe Roumagnac ^{2,3} and Pierre Lefeuvre ^{1,*}

¹ CIRAD, UMR PVBMT, F-97410 St Pierre, La Réunion, France; selim.ben_chehida@cirad.fr (S.B.C.); murielle.hoareau@cirad.fr (M.H.); jean-michel.lett@cirad.fr (J.-M.L.)

² CIRAD, PHIM, F-34398 Montpellier, France; Denis.Filloux@Cirad.Fr (D.F.);

emmanuel.fernandez@cirad.fr (E.F.); oumaima.moubset@cirad.fr (O.M.); charlotte.julian@cirad.fr (C.J.); laurence.blondin@cirad.fr (L.B.); philippe.roumagnac@cirad.fr (P.R.)

³ PHIM Plant Health Institute, University Montpellier, CIRAD, INRAE, Institut Agro, IRD, F-34398 Montpellier, France

* Correspondence: pierre.lefeuvre@cirad.fr

Abstract: Next-generation sequencing (NGS), through the implementation of metagenomic protocols, has led to the discovery of thousands of new viruses in the last decade. Nevertheless, these protocols are still laborious and costly to implement, and the technique has not yet become routine for everyday virus characterization. Within the context of CRESS DNA virus studies, we implemented two alternative long-read NGS protocols, one that is agnostic to the sequence (without a priori knowledge of the viral genome) and the other that use specific primers to target a virus (with a priori). Agnostic and specific long read NGS-based assembled genomes of two capulavirus strains were compared to those obtained using the gold standard technique of Sanger sequencing. Both protocols allowed the detection and accurate full genome characterization of both strains. Globally, the assembled genomes were very similar (99.5–99.7% identity) to the Sanger sequences consensus, but differences in the homopolymeric tracks of these sequences indicated a specific lack of accuracy of the long reads NGS approach that has yet to be improved. Nevertheless, the use of the bench-top sequencer has proven to be a credible alternative in the context of CRESS DNA virus study and could offer a new range of applications not previously accessible.

Keywords: MinION; nanopore sequencing; rolling circle amplification; viral metagenomics; CRESS DNA; capulavirus; homopolymer



Citation: Ben Chehida, S.; Filloux, D.; Fernandez, E.; Moubset, O.; Hoareau, M.; Julian, C.; Blondin, L.; Lett, J.-M.; Roumagnac, P.; Lefeuvre, P. Nanopore Sequencing Is a Credible Alternative to Recover Complete Genomes of Geminiviruses. *Microorganisms* **2021**, *9*, 903. <https://doi.org/10.3390/microorganisms9050903>

Academic Editor: Fred Asiegbu

Received: 24 March 2021

Accepted: 21 April 2021

Published: 23 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Recent advances in metagenomics applied to viruses has fostered a greater inventory of the viral diversity [1–4]. Hence, the large scale sampling of oceanic water [5,6], plants, animals, and humans [7–10], extreme environments [11], or the mining of genomic and transcriptomic data [12–15] have completely shifted our understanding of viral diversity and the function of viruses in host populations or even at the global ecosystem scale. However, these inventories remain largely incomplete, and the current knowledge of the virus diversity probably only represents the contour of the extant diversity [4]. The collection of hundreds of new genome sequences with, sometimes only remote resemblance to known viruses led to the acceptance of genome from metagenomic studies as genuine and legitimate taxonomic material for the description of new viruses [16–21], even without knowledge of the phenotype or the host associated to these viruses [17,22].

Yet, as access to metagenomics is usually costly and requires sophisticated technical expertise in data management and analysis [23,24]; for day-to-day analysis, classical Sanger sequencing remains more common [25]. Also, despite the potential of third generation sequencing technique to provide, in real time, hundreds of sequences with read length

of more than 15 kb [26,27], the resulting assembled genomes remains tainted with doubt. Indeed, assembly of the low base quality reads is required to obtain more accurate contigs but may result in chimeric genomes, or miss less frequent variants [17,28,29]. It would thus be beneficial to evaluate the use of third generation sequencing as an alternative to the gold standard of cloning and Sanger sequencing for everyday virus characterization. Indeed, several studies have successfully employed third-generation sequencing for virus detection or full genome recovery. It has been largely used to detect and sequence full genomes of a range of animal and human viruses, including influenza [30], Ebola [31], Dengue [32], Zika [33], or SRAS-Cov-2 [34]. Recently the method was also successfully applied to plant viruses for some yam infecting viruses [29], maize yellow mosaic virus [35] and plum pox virus [36].

Here, within the context of the study of circular replication-associated protein encoding single stranded (CRESS) DNA viruses (from the *Cressdnaviricota* phylum), we harness the power of third-generation sequencing for routine full genome assembly of viruses. A breakthrough in CRESS DNA virus studies was associated with the development of protocols using isothermal rolling circle amplification (RCA) in frequent association with enzymatic restriction, cloning and Sanger sequencing [37]. Metagenomic protocols using RCA were soon developed and greatly improved our knowledge of the CRESS DNA virus diversity [38,39]. Yet, the use of Illumina based protocols remains laborious and expensive for day-to-day viral genome characterization and sequences analysis usually requires the use of complex bioinformatics tools. The field would thus benefit from the development of a more convenient protocol on the MinION platform from the Oxford Nanopore Technologies (ONT). Indeed, recent studies have paved the way towards using nanopore sequencing, either from direct sequencing of total DNA extracts or after the application of RCA, to achieve full genome sequencing of CRESS-DNA viruses in general [40–42] or viruses from the *Geminiviridae* family in particular [43–45].

Two alternative protocols for use on the MinION bench-top sequencing device were designed in this study and the assemblies of the nanopore-sequenced reads of multiple strains of capulaviruses infecting a *Medicago arborea* plant [46] were compared to the Sanger genome sequences. *Capulavirus* is one of the nine genera of the *Geminiviridae* family [47]. This family is composed of plant-infecting viruses with genomes comprising one or two circular ssDNA of 2.5–5.2 kb encapsidated in twinned icosahedral (geminate) particles. They are transmitted by a high range of hemipterans (whiteflies, leafhoppers, aphids, and treehoppers) [48,49]. Whereas members of the *Geminivirus* family were first described in 1993 (ICTV), the standardization of RCA based protocols led to an explosion of its known diversity and the family regularly counts new genus-level lineage addition [47,50–52]. Following the description of the *Euphorbia caput-medusae* latent virus (EcmLV), the genus capulavirus has been proposed [53]. Their genome length ranged between 2.7 and 2.8 kb with two intergenic regions. The replication-associated protein (Rep) is expressed from a spliced complementary-strand transcript. A unique feature of capulavirus genomes is a complex arrangement of possible MP-encoding ORFs located in the 5' direction from the coat protein gene (*cp*) [47]. It is known to be transmitted by only two species of aphids: *Aphis Craccivora* [49] and *Dysaphis plantaginea* [54].

Our analysis revealed that MinION sequencing followed by read assembly results in genome sequences mostly similar to the consensus of the virus population that was obtained after Sanger sequencing of multiple isolates. The two alternate protocols used, one that does not require knowledge of the viruses present within the sample (without a priori) and the other designed to specifically amplify a given virus (with a priori), were successful for full genome assembly of the two capulavirus strains. MinION assembled consensus sequences present with more differences in the homopolymeric tracks but remain closely related to any sequence of the sample than any Sanger sequence to one another. Overall, both nanopore-based protocols are adapted to the genome size of the CRESS DNA viruses and the cost of the method is on par with the Sanger sequencing approach.

2. Material and Methods

2.1. Sampling and DNA Extraction

Leaf samples of an apparently asymptomatic *Medicago arborea* were collected in November 2019 at Montferrier-sur-Lez (France). The sample was stored at $-20\text{ }^{\circ}\text{C}$ before use. Total DNA was extracted using the DNeasy Plant DNA extraction kit (Qiagen, Hilden, Germany), following the manufacturer's instructions. DNA extract was stored at $-20\text{ }^{\circ}\text{C}$ before use. From a previous analysis [46], using a PCR amplification and Sanger sequencing, two strains of the capulavirus Trifolium virus 1 (TrV1-B and TrV1-C) were identified into the sample.

2.2. Full Genome Cloning and Sanger Sequencing

Pairs of abutting primers were designed to recover the full-length genome of TrV1-B and TrV1-C isolates. A two-step amplification was achieved, including a first amplification step using either the primer pair 3580F-CAPULUZARB-1F: 5'-ACT TGC CTG TCG CTC TAT CTT CTC CCT TGG AGA TGT AAT CTG CCA CGT CAG-3', and PR2-CAPULUZARB-2R: 5'-TTT CTG TTG GTG CTG ATA TTG CGG AGT TTT TGA GGA ACG AGG AAT ACT TAG AGC TTC A-3' for amplifying TrV1-B genomes or the primer pair 3580F-CAPUCORO-1F: 5'-ACT TGC CTG TCG CTC TAT CTT CAA CTG TCC TCC CTT TGC AAT GTA GTC AGC C-3' and PR2-CAPUCORO-2R: 5'-TTT CTG TTG GTG CTG ATA TTG CCG AGG AGC GAG GAC TTC TTA AGG CAA GT-3' for amplifying TrV1-C genomes. Amplification was carried out using the GoTaq[®] Master Mix Kit (Promega Corporation, Madison, WI, USA) and the following conditions: an initial denaturation at $95\text{ }^{\circ}\text{C}$ for 5 min, 8 cycles at $94\text{ }^{\circ}\text{C}$ for 30 s, $60\text{ }^{\circ}\text{C}$ for 30 s, $72\text{ }^{\circ}\text{C}$ for 3 min, and a final extension step at $72\text{ }^{\circ}\text{C}$ for 10 min. A common second amplification step was then performed using the primer pair (3580F: 5'-ACT TGC CTG TCG CTC TAT CTT C-3' and PR2: 5'-TTT CTG TTG GTG CTG ATA TTG C-3') and the GoTaq[®] Master Mix Kit (Promega Corporation, Madison, WI, USA). The amplification conditions were as followed: an initial denaturation at $95\text{ }^{\circ}\text{C}$ for 5 min, 25 cycles at $94\text{ }^{\circ}\text{C}$ for 30 s, $60\text{ }^{\circ}\text{C}$ for 30 s, $72\text{ }^{\circ}\text{C}$ for 3 min, and a final extension step at $72\text{ }^{\circ}\text{C}$ for 10 min. Amplification products of approximately 2.7–2.8 kb were gel purified, ligated to pGEM-T (Promega, Madison, WI, USA) and sequenced by standard Sanger sequencing using a primer walking approach.

2.3. Minion Sequencing

Two alternative protocols for MinION sequencing were used. In the first protocol, called hereafter the RCA-MinION protocol, a RCA amplification was performed using *Phi29* DNA polymerase (Illustra TempliPhi Amplification Kit, GE Healthcare, Chicago, IL, USA) by mixing 2 μL of total plant DNA extract with 5 μL of Sample Buffer before incubation at $95\text{ }^{\circ}\text{C}$ during 3 min. After cooling at room temperature, 0.2 μL of enzyme mix and 5 μL of Reaction Buffer were added before incubation at $30\text{ }^{\circ}\text{C}$ for 6 h followed by 20 min of polymerase deactivation at $65\text{ }^{\circ}\text{C}$. RCA products were cleaned-up using 2 \times of Sera-Mag Select Size Selection beads (GE Healthcare) and the 10 μL eluate were digested with 1 μL of T7 Endonuclease I (NEB), 2 μL of 5 \times buffer in a 10 μL reaction volume at $37\text{ }^{\circ}\text{C}$ during 1 h. The fragments were purified with a 1 \times Sera-Mag Select Size Selection beads and eluate with 10 μL of purified water. Library construction for MinION sequencing was performed using the PCR Barcoding Kit (SQK-PBK004), following the manufacturer's instructions but using SeraMag Select Size Selection beads (GE Healthcare) for DNA purification. Sequencing was performed as described below for the PCR-MinION procedure. Two Flongles (flow cell dongle) FLO-FLG001 were used for sequencing. Whereas in the first Flongle, a single RCA amplicon was sequenced, in the second, three distinct RCA amplicons were multiplexed.

In the second protocol, called hereafter the PCR-MinION protocol, a two-step amplification was carried out. In the first PCR, both sets of abutting primers (3580F-CAPULUZARB-1F/PR2-CAPULUZARB-2R and 3580F-CAPUCORO-1F/PR2-CAPUCORO-2R) described above for respectively amplifying the genomes of strains TrV1-B and TrV1-C and the GoTaq[®] Master Mix Kit (Promega Corporation, Madison WI, USA) were employed. The amplification conditions were as follows: an initial denaturation at 95 °C for 5 min, 15 cycles at 94 °C for 30 s, 60 °C for 30 s, 72 °C for 3 min, and a final extension step at 72 °C for 10 min. The second amplification step was performed using the cDNA Primer (cPRM) supplied in the PCR-cDNA Sequencing Kit (SQK-PCS109) and the LongAmp Taq 2X Master Mix Kit (New England Biolabs, Evry, France). The amplification conditions were as followed: an initial denaturation at 95 °C for 30 s, 20 cycles at 95 °C for 15 s, 62 °C for 15 s, 65 °C for 3 min, and a final extension step at 65 °C for 6 min. The amplicons were purified using Agencourt AMPure XP beads (Beckman Coulter, Brea, CA, USA) and MinION sequencing library was constructed using the PCR-cDNA Sequencing Kit (SQK-PCS109), following manufacturer's instructions. Sequencing was then performed on Flongle (FLO-FLG001) using MinKNOW software 19.06.8. Three Flongles were used, two for TrV1-B (Flongle 13 and Flongle 14) and another one for TrV1-C (Flongle 15).

2.4. MinION Sequence Assembly

For the reads obtained through the two MinION protocols, accurate basecalling was performed using Guppy (v4.09 or 4.015; [55]). Demultiplexing and adapter removal was then performed using Porechop v0.2.4 [56]. Quality of the reads was investigated using NanoPlot v1.33.0 [57].

The demultiplexed reads obtained from RCA-MinION procedure were assembled with FLYE 2.6 [58], using the "meta" and "plasmid" parameters and when possible circularized as monomers using a homemade script. Contigs were then subjected to a BLASTn search against a CRESS DNA reference sequence database obtained from GenBank. CRESS DNA contigs were then polished using Medaka v1.2.2 [59]. PCR-MinION sequences higher than 1500 nt for one run (TrV1-C) and 2000 nt for the two other run (TrV1-B), were assembled using Canu v1.8 [60]. CRESS DNA contigs were filtered using BLAST as described above. Contigs coverage was estimated after mapping the raw reads back to the assembled CRESS DNA sequences using Minimap2 v2.17 [61] (Figure 1).

2.5. Sequence Comparison and Phylogenetic Analysis

All the available capulavirus full genome sequences were downloaded from GenBank on 4 March 2021 and aligned with the sequences of strains TrV1-B and TrV1-C obtained after Sanger sequencing using MAFFT v7.475 [62]. A maximum likelihood (ML) phylogenetic tree was inferred using FastTree2 [63] using the "gtr" and "gamma" parameters. Branch supports were tested using SH-like local supports. Tree edition was performed using the ape R package [64]. In order to properly classify the sequences obtained, an analysis that include a subset of representative capulavirus from GenBank and the Sanger sequence obtained in this study was performed using SDT1.2 [65].

Sequences obtained using the three distinct protocols (i.e., Sanger, PCR-MinION, and RCA-MinION procedures) were aligned together with MAFFT v7.475 before manual edit of the alignment. A home-made R script was used for sequence comparison and mutation count. Mutations were classified in three categories: substitution, insertion/deletion (INDEL) and homopolymer length variation (HLV) (Figure 1). ML trees were inferred from these alignments using FastTree2 as described above.

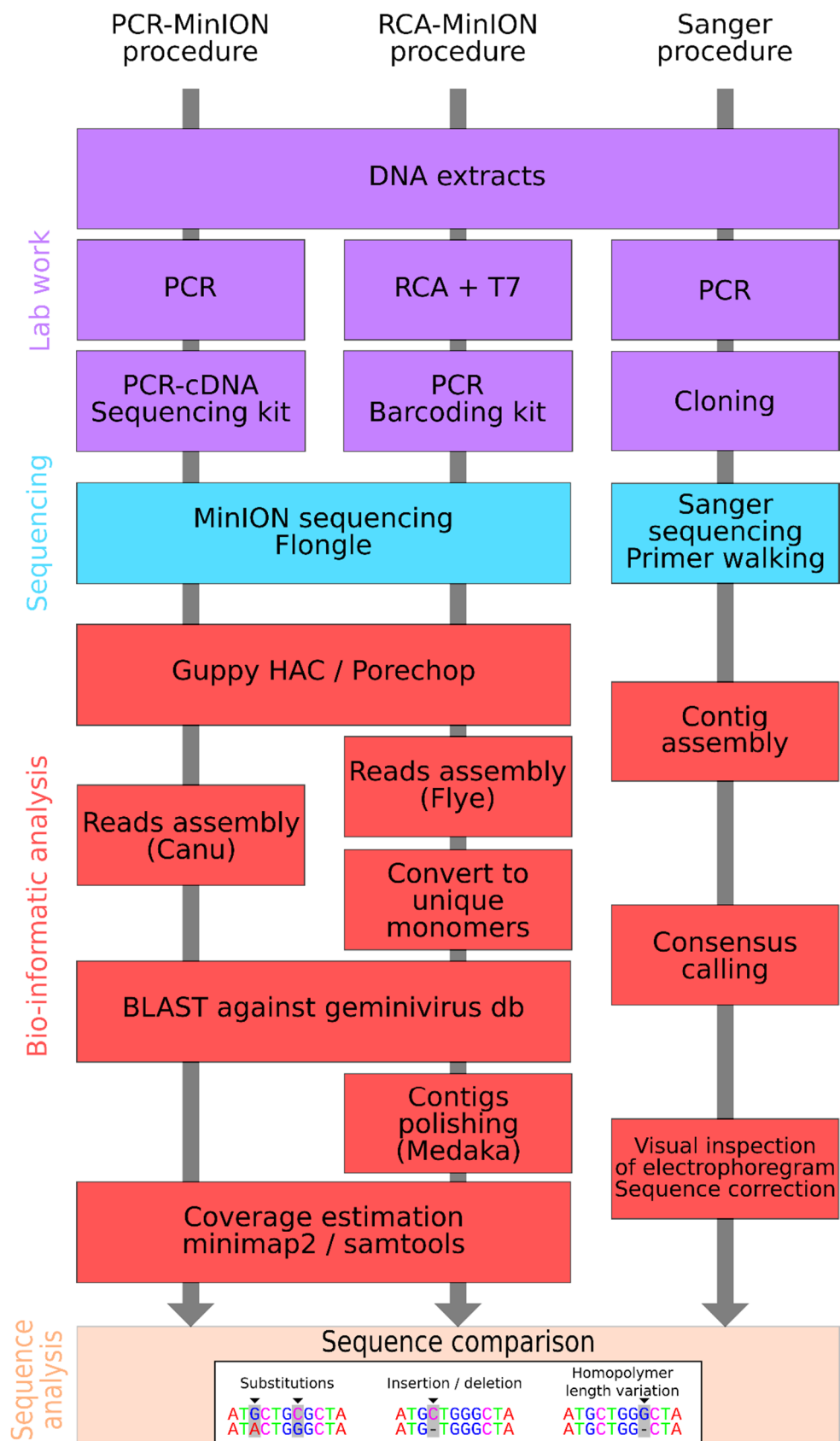


Figure 1. Schematic representation of the three distinct sequencing methodologies used including the molecular biology procedure (purple boxes) the sequencing procedure (light blue boxes), the bio-informatics procedure (red boxes) and the method comparison (orange boxes).

3. Results and Discussion

3.1. Sanger Sequences

A total of 46 sequences were obtained after cloning and Sanger sequencing. These sequence groups in two distinct clades (Figure S1) that share a mean identity of 90.3%. Within the first clade ($n = 19$), sequences present with identity ranging from 99.5 to 100% with each other and were most similar to the isolate BG2_capuz_47 of the B strain of TrV1 (GenBank accession number MW698819) with a minimum identity of 99.7%. Within the second clade ($n = 27$), sequences present identity ranging from 99.0 to 100% with each other and were most closely related to the isolate BG2_coro_02-2 of the C strain of TrV1 (GenBank accession number MW698820) with a minimum identity of 99.1%. Therefore, our two groups of sequences belong to two distinct strains (TrV1-B and TrV1-C) of the capulavirus *Trifolium virus 1* species (Figure S1). It is important to notice here that among the 46 isolates, five of the TrV1-B isolates and seven of the TrV1-C isolates present with defective genomes. Two of the TrV1-B defective isolates have deletions that encompass a fraction of the V3 gene, two other isolates have deletions within the gene of the replication-associated protein and one last has a deletion that encompasses the *rep* gene. Four sequences of the TrV1-C isolates have deletions that encompass a fraction of the CP gene. Among these, the deletions also span a fraction of the V3 gene or the *rep* gene, depending on the isolate. One isolate has a deletion that encompasses a fraction of all the gene encoded in the complement strand. These twelve defective sequences were excluded from downstream analysis. The genomic organization of the remaining sequences confirms the presence of a short intergenic region (SIR) and a long intergenic region (LIR), a characteristic inverted repeat potentially capable of forming a stem-loop structure that included a conserved nonanucleotide sequence TAATATTAC present at almost all geminivirus virion-strand replication origins, the *cp*, a spliced complementary-strand intron-containing transcript which expresses replication-associated protein gene (*rep*), a large complementary-sense ORF (C3) that is completely embedded within the *rep* gene, and a complex arrangement of possible MP-encoding ORFs located in the 5' direction from the *cp* gene, which is a unique feature of *Capulavirus* genomes [47]. Four and three sequences of the TrV1-B and -C strains, present truncated ORFs. The full genome sequences of the isolate without ORFs truncation are available on GenBank under the accession numbers MW698819–MW698821 and MW768713–MW768736.

3.2. Long Read Sequencing and Assembly

The RCA-MinION generated reads that confirmed the presence of both TrV1-B and TrV1-C strains in the *Medicago arborea* sample (Figure 2). The raw sequencing statistics are available in Table 1. From Flongle 1 and Flongle 2, 188,123 and 273,088 raw reads were obtained from which 110,830 (59%) and 152,076 (56%) barcoded reads passed the quality control (Figure S2), respectively. The median read length was 1154 bp with a longest read of 10,491 bp. From Flongle 1, only 27 reads (0.02%) mapped with the capulavirus references. From Flongle 2, barcodes were retrieved from 65,413 reads, 40,242 reads and 46,421 reads for each of the three barcodes, respectively. From 6 to 8.5% of those reads mapped with the capulavirus sequences. Although it was performed on the same DNA extract, RCA amplification yielded more than two order of magnitude less viral sequence for the Flongle 1 amplification than those performed for Flongle 2. It highlights known bias of the RCA amplification [66,67] that were already evidenced in the context of CRESS virus amplification [68]. All the four distinct sequence sets (one barcode sample from Flongle 1 and three for Flongle 2) were then submitted to the assembly and circularization pipeline. For each of the four barcodes, unique contigs corresponding to the full genome sequences of the two strains were obtained. The four TrV1-B sequence lengths ranged from 2745 to 2769 bp and were at least 99.5% similar to any Sanger sequence. One of the TrV1-B sequences (RCA-Minion_10_2, Flongle 2 barcode 10, Table 1) present with a region that seems to be mis-assembled (100 nt in length, see grey tracks in Figure 3A). The four TrV1-C sequences length ranged from 2763 to 2771 bp and were at least 99.1% similar to any

Sanger sequence. Again, a probable mis-assembly (68 nt in length, grey tracks in Figure 3B) was present within one sequence (RCA-MinION_01_1). As none of the raw minion reads supported the presence of this putative recombinant region, it has not been taken into consideration for further analyses.

Table 1. Long read sequencing statistics.

Flongle Id	Raw Reads	Passed Reads	Barcode ID	Trimmed Reads	Capulavirus Reads	TrV1-B Reads	TrV1-C Reads	Length TrV1-B Assembly	Length TrV1-C Assembly
1	188,123	162,263	1	110,830	27	14	13	2745	2771
			10	65,413	4809	3230	1579	2769	2764
2	273,088	215,143	11	40,242	3423	2244	1179	2748	2763
			12	46,421	2803	1585	1218	2746	2765
13	492,922	386,099	-	385,029	380,933	380,933	-	2731	-
14	414,665	371,700	-	370,755	367,381	367,381	-	2739	-
15	768,144	714,730	-	602,579	337,338	-	337,338	-	2754

The PCR-MinION generated reads that also confirmed the presence of TrV1-B and TrV1-C strains after RCA-MinION and Sanger sequencing. From Flongle 13, Flongle 14, and Flongle 15, 492,922, 414,665, and 768,144 raw reads were obtained from which 386,099 (78.3%), 371,700 (89.6%), and 714,730 (93.0%) reads passed the quality control (Table 1, Figure S2). The median length of the passed reads was 643 bp with a longest read of 9816 bp. From Flongle 13, 14 and 15, 380,933 reads (98.7%), 367,381 reads (98.8%) and 337,338 reads (47.2%) mapped with the TrV1-B and -C reference sequences. All the three read sets were then submitted to the assembly. For every Flongle, contigs corresponding to the full genome sequence of the TrV1-B and TrV1-C Sanger references were obtained. The two TrV1-B sequences length ranged from 2731 to 2739 bp and were at least 99.7% similar to any Sanger sequence. The TrV1-C sequence is 2754 bp length and were at least 99.3% similar to any Sanger sequence (Figure 2).

3.3. Sequences Comparison

In order to more precisely determine the nature of the differences between the sequences obtained through the different procedures, per capulavirus strains, all the full genome sequences obtained were compared to the consensus of the Sanger sequences. With more than 99.1% nucleotide identity for both the RCA-MinION and PCR-MinION sequences, the two methods demonstrate their ability to recover full genomes that are accurately assigned to strains TrV1-B and -C and whose sequence are representative of the viral population they originate. However, none of the assemblies obtained from MinION sequences was 100% identical to the Sanger sequence.

Beside mis-assemblies (grey tracks on Figure 3), the differences between sequences were classified in three categories with substitution (green ticks on Figure 3), INDEL (blue ticks) and HLV (red ticks). It must be noticed here that HLVs are a category of INDELs but were count separately as HLVs are recognized as the main source of errors in nanopore sequencing [69–72].

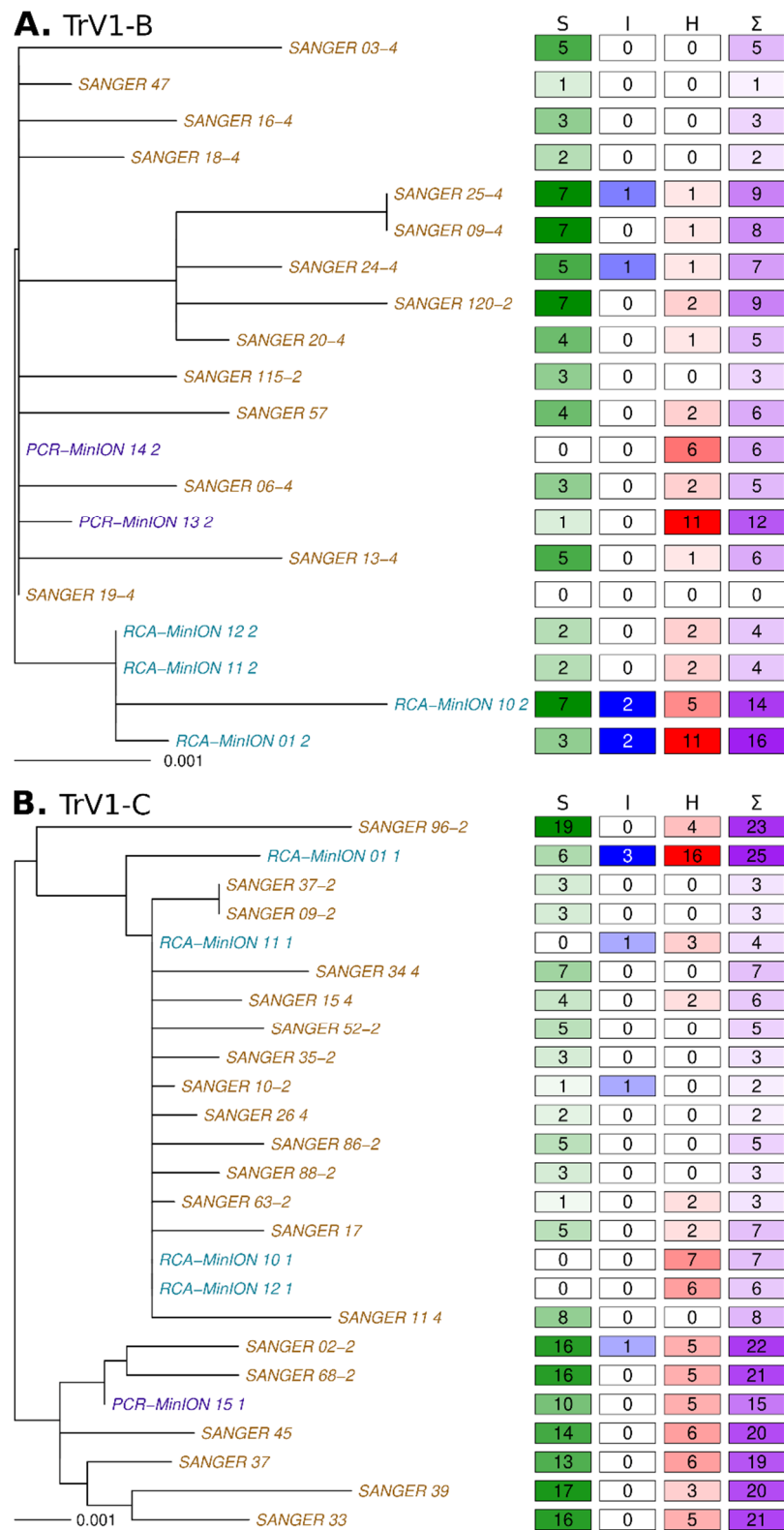


Figure 2. Maximum likelihood phylogenetic tree of all the sequences of the TrV1-B (A) and TrV1-C (B) strains obtained after Sanger sequencing (brown tips) or MinION sequencing followed by read assemblies (blue and purple tips for the RCA-MinION and PCR-MinION procedures respectively) on the left along with a matrix presenting the number of mutation relative to the Sanger consensus on the right. The numbers in the four columns present the substitutions (S), INDELs (I), HLVs (H) and the sum (Σ) of all these variations.

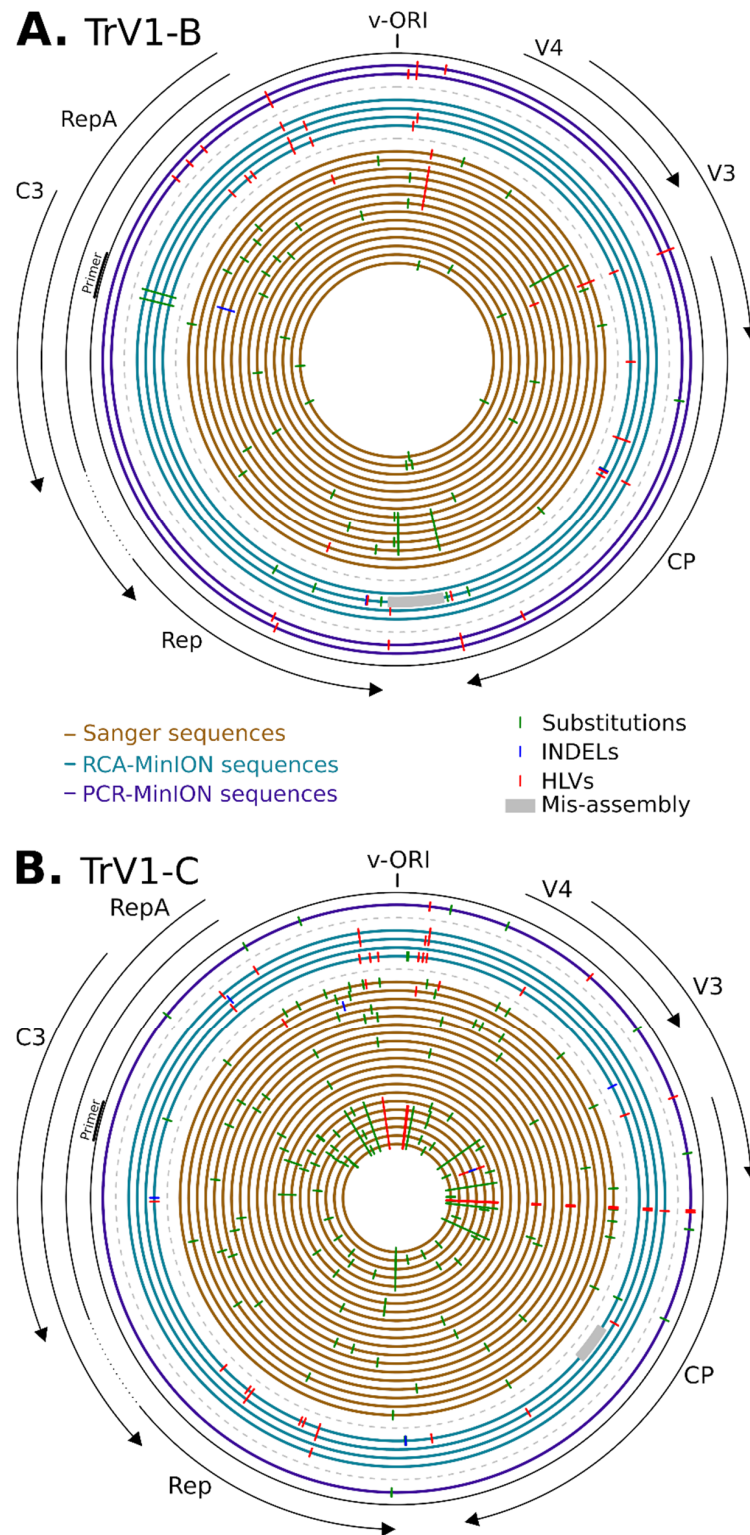


Figure 3. Diagram presenting the positions of the mutations along the genome of the TrV1-B (A) and TrV1-C (B) isolates and assemblies. Concentric circles represent each complete genome sequence obtained, with from the center to the outer, Sanger sequences in brown, RCA-MinION in blue and PCR-MinION in purple. Substitutions, INDELs and homopolymer length variations are represented with green, blue and red ticks respectively. Grey areas represent large deletions. Positions are relative to reference sequences MW698819 for TrV1-B and MW698820 for TrV1-C. The origin of replication (v-ORI) is indicated on top and the ORFs are represented on the outside of the figure.

First, Sanger sequences differ from each other (and to the consensus) mostly with substitutions (accounting for 79.5% of the variations) and more marginally with HLVs (19.4% of the variations) and INDELs (1.5% of the variations) (Figure 2). Except for two of the TrV1-C sequences, all the other Sanger sequences were unique and the number of variations between all these sequences was up to 12 and 27 for TrV1-B and TrV1-C, respectively. A total of 46 and 113 polymorphic sites were present in the Sanger sequences for TrV1-B and TrV1-C, respectively (Figure 3). Whereas these polymorphic sites tend to be more frequently present within non-coding regions (binomial test p -value of 0.090 and 0.016 for TrV1-B and TrV1-C, respectively), these sites had globally more mutations among all the Sanger sequences (binomial test p -value of 3.7×10^{-3} and 7.7×10^{-6} for TrV1-B and TrV1-C, respectively). Whereas sequencing errors can explain a fraction of the mutation, with the high accuracy of Sanger sequencing in mind (per base Phred quality score of 50 [73]), one can expect that most of the variations uncovered during the analysis are genuine and represent the biological variation associated with the diversity of the viral population infecting the plant [74–76].

The analysis of the RCA-MinION and PCR-MinION sequences revealed a different pattern of polymorphism. Two sequences, one of TrV1-B and one of TrV1-C, both assembled from a restricted number of reads (Flongle 1, RCA-MinION_01-1 and RCA-MinION_01-2) were, as expected, less accurate. Coverage of the assembly, (e.g., representing the mean number of times every position of an assembly was read) is therefore a good indicator of the reliability of the resulting assembly (Figure 4). Among the nine other MinION sequences assembled, seven present with no substitution to the consensus and one with ten substitutions (Figure 2). Through the TrV1-B sequences there were few common mutations: only a single HLV was common between Sanger and MinION assemblies (Figure 3A). PCR-MinION and RCA-MinION sequences presented four common HLVs (Figure 3A). For TrV1-C, Sanger sequences presented one substitution, no INDEL and three HLVs in common with MinION sequences (Figure 3B). PCR-MinION and RCA-MinION sequences presented one common substitution, three common HLVs and no common INDELs (Figure 3B). As these sequences were obtained after the assembly of reads, we were unable to catch the diversity of the distinct variants forming the viral population but rather to obtain a sequence very similar to the consensus of that population. Conversely to the reduced number of substitutions in comparison to the consensus, the assemblies present with larger numbers of INDELs (from 1 to 3) and HLVs (2 to 16). Some of these variations were also found in the Sanger sequences and it is probable that the assemblies actually represented some of the variability within the population. Nevertheless, for six INDELs and 38 HLVs, no corresponding mutations were found, and most would induce frameshift or protein truncation (Figure 3). Despite the use of a dedicated sequence correction program, multiple HLVs remained in the assembled sequences.

3.4. Defective Genome and Sequence Coverage

Defective genomes are frequently detected within geminivirus populations [77]. For instance, twelve of the Sanger sequences displayed large deletions in comparison to full reference genomes. As MinION raw sequences are obtained after direct reads of either PCR or RCA amplicons, they should capture the diversity of defective genomes of the viral populations. Indeed, to obtain full genome assemblies using the PCR-MinION procedure, a selection of the reads approaching full genome size was required. The analysis of coverage of the reads (Figure 4) confirms the pervasive nature of defective genomes for both the TrV1-B and -C strains. The highest coverage was obtained for regions encompassing the stem-loop and reads were frequently missing most of the coding regions (see the blue lines for RCA-MinION). Importantly, it must be noticed here that the coverage inferred using the PCR-MinION procedure are not representative of the global population but rather represent the subset of virus that contains the priming site of the abutting primers used in the PCR (indicated with the verticals red dotted lines in Figure 4). The results indicate that every position of the TrV1-B strains present with a high coverage, most of the amplicons of

TrV1-C were defective for a region encompassing the whole CP gene (from position 147 to 1559).

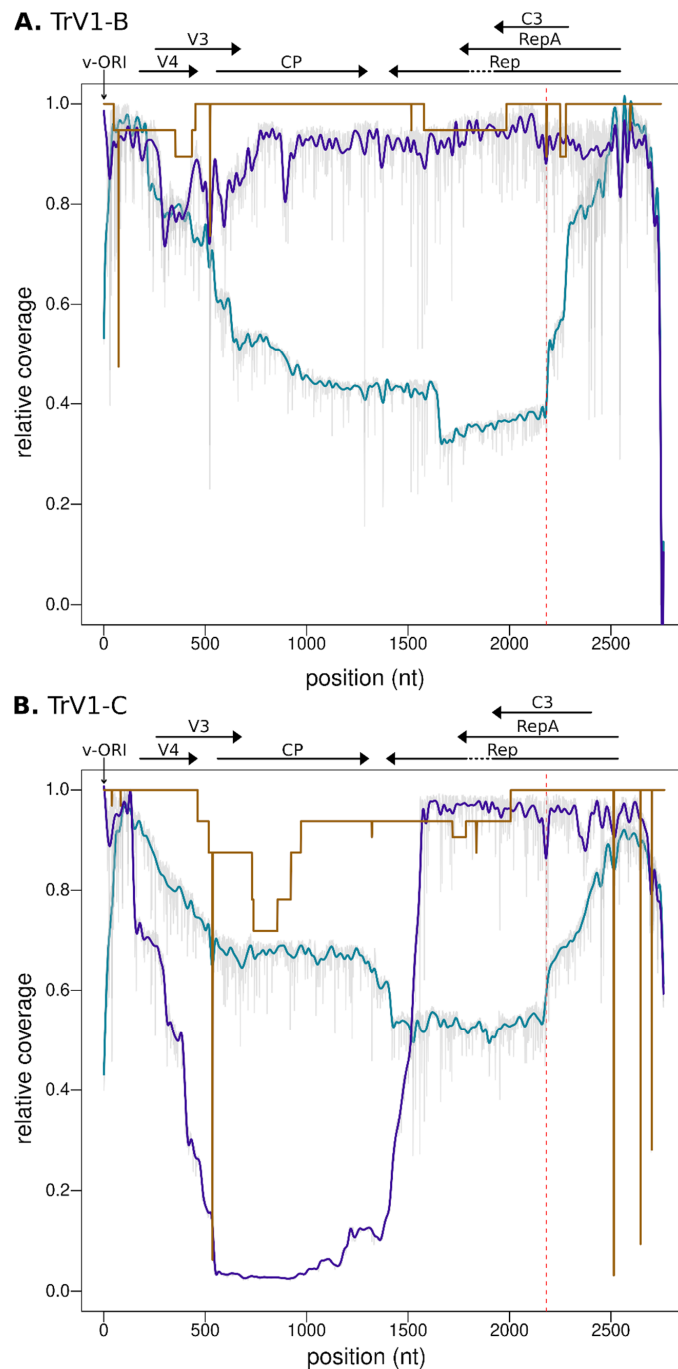


Figure 4. Coverage plot along the reference genome of TrV1-B (A) and TrV1-C (B). The coverage, interpreted here as the fold coverage of a given position divided by the maximum coverage of the whole genome, is represented with a brown line for Sanger sequences, a blue line for RCA-MinION sequences and a purple line for PCR-MinION sequences. MinION coverage are interpolations of the raw coverages presented here in grey on the background. The position of the abutting primers used in the PCR-MinION procedure is indicated by the vertical red dotted line. Positions are relative to the reference genome MW698819 for TrV1-B and MW698820 for TrV1-C. ORFs of these reference genomes are symbolized with horizontal arrows on top of the figure. The Origin of replication is indicated with an arrow. The position 1 of each sequence correspond to the nick-site within the stem loop structure.

4. Conclusions

From an asymptomatic sample of *Medicago arborea*, two distinct strains of the capulavirus TrV1 have been cloned and sequenced by the Sanger methodology. Using both the *a priori* and the agnostic nanopore-based procedures, both TrV1-B and TrV1-C strains were detected, and full genome sequences were assembled. Despite being very similar to the consensus of Sanger sequences, mutations specific to the MinION assemblies were detected mostly within homopolymeric regions of the genomes, which is in agreement with other studies that have also pinpointed higher number of errors associated with homopolymer lengths [78,79]. Whereas it could be argued that the HLV errors would be reduced with the increasing accuracy of sequencing, the development of a new base caller technologies and correction algorithm [80], current MinION sequences assemblies may be avoided for some specific applications where the exact nucleotide sequence is required. Otherwise, for other applications, such as virus discovery, virus classification, or recombination analysis, MinION assemblies represent a competitive alternative to Sanger sequencing. Although, similarly to other NGS protocols, MinION based studies require the use of sophisticated bioinformatics tools for data management and analysis and despite the specific drawbacks on sequence quality, bench top sequencer such as the MinION would probably become routinely used in the laboratory. It allows a real-time detection and diagnostic of multiple viral strains or virus species in a single run. For niche applications, such as the exploration of geminivirus genomes, which do not exceed 10 kb, nanopore sequencing is poised to push the cost and performance limits of sequencing technologies. The high reactivity offered by the platform would make it more and more democratized as a mobile real-time plant disease diagnostic tool.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/microorganisms9050903/s1>. Figure S1: Maximum likelihood phylogenetic tree of all the *Capulavirus* genus full genome sequences obtained from NCBI along with the two TrV1-B and TrV1-C genotypes obtained in this study; Supplementary Figure S2: Reads length (*x*-axis) against quality score (*y*-axis) for all passed reads (A and C) and capulavirus only passed reads (B and D) obtained using the RCA-MinION and PCR-MinION procedures.

Author Contributions: D.F., P.L. and P.R. conceived and designed the experiments. S.B.C., L.B., E.F., M.H., C.J. and O.M. performed the experiments. S.B.C., D.F. and P.L. analyzed and interpreted the data. S.B.C. and P.L. wrote original draft. D.F., P.L., J.-M.L. and P.R. revised and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the European Union (ERDF, contract GURDT I2016-1731-0006632), the *Conseil Régional de la Réunion* and CIRAD. SBC is a recipient of a PhD fellowship from CIRAD and ANR (Phytovirus project number: ANR-19-CE35-0008-02).

Data Availability Statement: Publicly available datasets were analyzed in this study. Sanger sequences accession numbers [MW698819–MW698821, MW768713–MW768736] are available on NCBI Genbank. MinION data are available at the NCBI Short read archive under the BioProject [PRJNA715304].

Acknowledgments: This work was conducted on the Plant Protection Platform (3P, IBISA).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dolja, V.V.; Koonin, E.V. Metagenomics reshapes the concepts of RNA virus evolution by revealing extensive horizontal virus transfer. *Virus Res.* **2018**, *244*, 36–52. [[CrossRef](#)] [[PubMed](#)]
2. Greninger, A.L. A decade of RNA virus metagenomics is (not) enough. *Virus Res.* **2018**, *244*, 218–229. [[CrossRef](#)]
3. Koonin, E.V.; Dolja, V.V. Metaviromics: A tectonic shift in understanding virus evolution. *Virus Res.* **2018**, *246*, A1–A3. [[CrossRef](#)] [[PubMed](#)]
4. Roux, S.; Matthijnsens, J.; Dutilh, B.E. Metagenomics in Virology. *Encycl. Virol.* **2021**, *1*, 133–140. [[CrossRef](#)]
5. Roux, S.; Brum, J.R.; Dutilh, B.E.; Sunagawa, S.; Duhaime, M.B.; Loy, A.; Poulos, B.T.; Solonenko, N.; Lara, E.; Poulain, J.; et al. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nat. Cell Biol.* **2016**, *537*, 689–693. [[CrossRef](#)]

6. McMullen, A.; Martinez-Hernandez, F.; Martinez-Garcia, M. Absolute quantification of infecting viral particles by chip-based digital polymerase chain reaction. *Environ. Microbiol. Rep.* **2019**, *11*, 855–860. [[CrossRef](#)]
7. Breitbart, M.; Hewson, I.; Felts, B.; Mahaffy, J.M.; Nulton, J.; Salamon, P.; Rohwer, F. Metagenomic Analyses of an Uncultured Viral Community from Human Feces. *J. Bacteriol.* **2003**, *185*, 6220–6223. [[CrossRef](#)]
8. Wylie, K.M.; Weinstock, G.M.; Storch, G.A. Emerging view of the human virome. *Transl. Res.* **2012**, *160*, 283–290. [[CrossRef](#)] [[PubMed](#)]
9. Lecuit, M.; Eloit, M. The diagnosis of infectious diseases by whole genome next generation sequencing: A new era is opening. *Front. Cell. Infect. Microbiol.* **2014**, *4*, 25. [[CrossRef](#)] [[PubMed](#)]
10. Maclot, F.; Candresse, T.; Filloux, D.; Malmstrom, C.M.; Roumagnac, P.; Van Der Vlugt, R.; Massart, S. Illuminating an Ecological Blackbox: Using High Throughput Sequencing to Characterize the Plant Virome Across Scales. *Front. Microbiol.* **2020**, *11*, 578064. [[CrossRef](#)]
11. Dávila-Ramos, S.; Castelán-Sánchez, H.G.; Martínez-Ávila, L.; Sánchez-Carbente, M.D.R.; Peralta, R.; Hernández-Mendoza, A.; Dobson, A.D.W.; Gonzalez, R.A.; Pastor, N.; Batista-García, R.A. A Review on Viral Metagenomics in Extreme Environments. *Front. Microbiol.* **2019**, *10*, 2403. [[CrossRef](#)] [[PubMed](#)]
12. Goodwin, S.; McPherson, J.D.; McCombie, W.R. Coming of age: Ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **2016**, *17*, 333–351. [[CrossRef](#)]
13. Shi, M.; Lin, X.-D.; Tian, J.-H.; Chen, L.-J.; Chen, X.; Li, C.-X.; Qin, X.-C.; Li, J.; Cao, J.-P.; Eden, J.-S.; et al. Redefining the invertebrate RNA virosphere. *Nat. Cell Biol.* **2016**, *540*, 539–543. [[CrossRef](#)]
14. Shi, M.; Zhang, Y.-Z.; Holmes, E.C. Meta-transcriptomics and the evolutionary biology of RNA viruses. *Virus Res.* **2018**, *243*, 83–90. [[CrossRef](#)]
15. Zhang, Y.-Z.; Shi, M.; Holmes, E.C. Using Metagenomics to Characterize an Expanding Virosphere. *Cell* **2018**, *172*, 1168–1172. [[CrossRef](#)] [[PubMed](#)]
16. Rosario, K.; Duffy, S.; Breitbart, M. A field guide to eukaryotic circular single-stranded DNA viruses: Insights gained from metagenomics. *Arch. Virol.* **2012**, *157*, 1851–1871. [[CrossRef](#)] [[PubMed](#)]
17. Simmonds, P.; Adams, M.J.; Benkő, M.; Breitbart, M.; Brister, J.R.; Carstens, E.B.; Davison, A.J.; Delwart, E.; Gorbalenya, A.E.; Harrach, M.B.B.; et al. Virus taxonomy in the age of metagenomics. *Nat. Rev. Genet.* **2017**, *15*, 161–168. [[CrossRef](#)]
18. Batovska, J.; Lynch, S.E.; Rodoni, B.C.; Sawbridge, T.I.; Cogan, N.O. Metagenomic arbovirus detection using MinION nanopore sequencing. *J. Virol. Methods* **2017**, *249*, 79–84. [[CrossRef](#)] [[PubMed](#)]
19. Roux, S.; Adriaenssens, E.M.; Dutilh, B.E.; Koonin, E.V.; Kropinski, A.M.; Krupovic, M.; Kuhn, J.H.; Lavigne, R.; Brister, J.R.; Varsani, A.; et al. Minimum Information about an Uncultivated Virus Genome (MIUViG). *Nat. Biotechnol.* **2019**, *37*, 29–37. [[CrossRef](#)] [[PubMed](#)]
20. Claverie, S.; Ouattara, A.; Hoareau, M.; Filloux, D.; Varsani, A.; Roumagnac, P.; Martin, D.P.; Lett, J.-M.; Lefeuvre, P. Exploring the diversity of Poaceae-infecting mastreviruses on Reunion Island using a viral metagenomics-based approach. *Sci. Rep.* **2019**, *9*, 1–11. [[CrossRef](#)]
21. International Committee on Taxonomy of Viruses Executive Committee. The new scope of virus taxonomy: Partitioning the virosphere into 15 hierarchical ranks. *Nat. Microbiol.* **2020**, *5*, 668–674. [[CrossRef](#)]
22. Roossinck, M.J.; Martin, D.P.; Roumagnac, P. Plant Virus Metagenomics: Advances in Virus Discovery. *Phytopathology* **2015**, *105*, 716–727. [[CrossRef](#)]
23. Quiñones-Mateu, M.E.; Avila, S.; Reyes-Teran, G.; Martinez, M.A. Deep sequencing: Becoming a critical tool in clinical virology. *J. Clin. Virol.* **2014**, *61*, 9–19. [[CrossRef](#)] [[PubMed](#)]
24. Mikheyev, A.S.; Tin, M.M.Y. A first look at the Oxford Nanopore MinION sequencer. *Mol. Ecol. Resour.* **2014**, *14*, 1097–1102. [[CrossRef](#)]
25. Sanger, F.; Coulson, A. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* **1975**, *94*, 441–448. [[CrossRef](#)]
26. Jain, M.; Olsen, H.E.; Paten, B.; Akeson, M. The Oxford Nanopore MinION: Delivery of nanopore sequencing to the genomics community. *Genome Biol.* **2016**, *17*, 1–11. [[CrossRef](#)]
27. Pomerantz, A.; Peñafiel, N.; Arteaga, A.; Bustamante, L.; Pichardo, F.; Coloma, L.A.; Barrio-Amorós, C.L.; Salazar-Valenzuela, D.; Prost, S. Real-time DNA barcoding in a rainforest using nanopore sequencing: Opportunities for rapid biodiversity assessments and local capacity building. *GigaScience* **2018**, *7*, 1–14. [[CrossRef](#)] [[PubMed](#)]
28. Xu, Y.; Lewandowski, K.; Lumley, S.; Pullan, S.; Vipond, R.; Carroll, M.; Foster, D.; Matthews, P.C.; Peto, T.; Crook, D. Detection of Viral Pathogens with Multiplex Nanopore MinION Sequencing: Be Careful With Cross-Talk. *Front. Microbiol.* **2018**, *9*, 2225. [[CrossRef](#)]
29. Filloux, D.; Fernandez, E.; Loire, E.; Claude, L.; Galzi, S.; Candresse, T.; Winter, S.; Jeeva, M.L.; Makesh Kumar, T.; Martin, D.P.; et al. Nanopore-based detection and characterization of yam viruses. *Sci. Rep.* **2018**, *8*, 1–11. [[CrossRef](#)] [[PubMed](#)]
30. Ewang, J.; Moore, N.E.; Edeng, Y.-M.; Eccles, D.A.; Hall, R.J. MinION nanopore sequencing of an influenza genome. *Front. Microbiol.* **2015**, *6*, 766. [[CrossRef](#)]
31. Quick, J.; Loman, N.J.; Duraffour, S.; Simpson, J.T.; Severi, E.; Cowley, L.; Bore, J.A.; Koundouno, R.; Dudas, G.; Mikhail, A.; et al. Real-time, portable genome sequencing for Ebola surveillance. *Nat. Cell Biol.* **2016**, *530*, 228–232. [[CrossRef](#)]

32. Yamagishi, J.; Runtuwene, L.R.; Hayashida, K.; Mongan, A.E.; Thi, L.A.N.; Thuy, L.N.; Nhat, C.N.; Limkittikul, K.; Sirivichayakul, C.; Sathirapongsasuti, N.; et al. Serotyping dengue virus with isothermal amplification and a portable sequencer. *Sci. Rep.* **2017**, *7*, 1–10. [[CrossRef](#)]
33. Quick, J.; Grubaugh, N.D.; Pullan, S.T.; Claro, I.M.; Smith, A.D.; Gangavarapu, K.; Oliveira, G.; Robles-Sikisaka, R.; Rogers, T.F.; Beutler, N.A.; et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protoc.* **2017**, *12*, 1261–1276. [[CrossRef](#)] [[PubMed](#)]
34. Kim, D.; Lee, J.-Y.; Yang, J.-S.; Kim, J.W.; Kim, V.N.; Chang, H. The Architecture of SARS-CoV-2 Transcriptome. *Cell* **2020**, *181*, 914–921.e10. [[CrossRef](#)]
35. Adams, I.P.; Braidwood, L.A.; Stomeo, F.; Phiri, N.; Uwumukiza, B.; Feyissa, B.; Mahuku, G.; Wangi, A.; Smith, J.; Mumford, R.; et al. Characterising Maize Viruses Associated with Maize Lethal Necrosis Symptoms in Sub Saharan Africa. *bioRxiv* **2017**. preprint. [[CrossRef](#)]
36. Badial, A.B.; Sherman, D.; Stone, A.; Gopakumar, A.; Wilson, V.; Schneider, W.; King, J. Nanopore Sequencing as a Surveillance Tool for Plant Pathogens in Plant and Insect Tissues. *Plant Dis.* **2018**, *102*, 1648–1652. [[CrossRef](#)] [[PubMed](#)]
37. Inoue-Nagata, A.K.; Albuquerque, L.C.; Rocha, W.B.; Nagata, T. A simple method for cloning the complete begomovirus genome using the bacteriophage ϕ 29 DNA polymerase. *J. Virol. Methods* **2004**, *116*, 209–211. [[CrossRef](#)] [[PubMed](#)]
38. Rosario, K.; Dayaram, A.; Marinov, M.; Ware, J.; Kraberger, S.; Stainton, D.; Breitbart, M.; Varsani, A. Diverse circular ssDNA viruses discovered in dragonflies (Odonata: Epiprocta). *J. Gen. Virol.* **2012**, *93*, 2668–2681. [[CrossRef](#)]
39. Rosario, K.; Schenck, R.O.; Harbeitner, R.C.; Lawler, S.N.; Breitbart, M. Novel circular single-stranded DNA viruses identified in marine invertebrates reveal high sequence diversity and consistent predicted intrinsic disorder patterns within putative structural proteins. *Front. Microbiol.* **2015**, *6*, 696. [[CrossRef](#)]
40. Theuns, S.; Vanmechelen, B.; Bernaert, Q.; Deboutte, W.; Vandenhole, M.; Beller, L.; Matthijnsens, J.; Maes, P.; Nauwynck, H.J. Nanopore sequencing as a revolutionary diagnostic tool for porcine viral enteric disease complexes identifies porcine kobovirus as an important enteric virus. *Sci. Rep.* **2018**, *8*, 1–13. [[CrossRef](#)]
41. Boykin, L.M.; Sseruwagi, P.; Alicai, T.; Ateka, E.; Mohammed, I.U.; Stanton, J.-A.L.; Kayuki, C.; Mark, D.; Fute, T.; Erasto, J.; et al. Tree Lab: Portable genomics for Early Detection of Plant Viruses and Pests in Sub-Saharan Africa. *Genes* **2019**, *10*, 632. [[CrossRef](#)] [[PubMed](#)]
42. Cao, J.; Zhang, Y.; Dai, M.; Xu, J.; Chen, L.; Zhang, F.; Zhao, N.; Wang, J. Profiling of Human Gut Virome with Oxford Nanopore Technology. *Med. Microecol.* **2020**, *4*, 100012. [[CrossRef](#)]
43. Chalupowicz, L.; Dombrowsky, A.; Gaba, V.; Luria, N.; Reuven, M.; Beerman, A.; Lachman, O.; Dror, O.; Nissan, G.; Manulis-Sasson, S. Diagnosis of plant diseases using the Nanopore sequencing platform. *Plant Pathol.* **2018**, *68*, 229–238. [[CrossRef](#)]
44. Naito, F.Y.B.; Melo, F.L.; Fonseca, M.E.N.; Santos, C.A.F.; Chanes, C.R.; Ribeiro, B.M.; Gilbertson, R.L.; Boiteux, L.S.; Pereira-Carvalho, R.D.C. Nanopore sequencing of a novel bipartite New World begomovirus infecting cowpea. *Arch. Virol.* **2019**, *164*, 1907–1910. [[CrossRef](#)]
45. Leiva, A.M.; Siriwan, W.; Alvarez, D.L.; Barrantes, I.; Hemniam, N.; Saokham, K.; Cuellar, W.J. Nanopore-Based Complete Genome Sequence of a Sri Lankan Cassava Mosaic Virus (Geminivirus) Strain from Thailand. *Microbiol. Resour. Announc.* **2020**, *9*, 9. [[CrossRef](#)]
46. Ma, Y.; Svanella-Dumas, L.; Julian, C.; Galzi, S.; Fernandez, E.; Yvon, M.; Pirolles, E.; Lefebvre, M.; Filloux, D.; Roumagnac, P.; et al. Genome characterization and diversity of trifolium virus 1: Identification of a novel legume-infective capulavirus. *Arch. Virol.* **2021**, in press.
47. Varsani, A.; Roumagnac, P.; Fuchs, M.; Navas-Castillo, J.; Moriones, E.; Idris, A.; Briddon, R.W.; Rivera-Bustamante, R.; Zerbini, F.M.; Martin, D.P. Capulavirus and Grabovirus: Two new genera in the family Geminiviridae. *Arch. Virol.* **2017**, *162*, 1819–1831. [[CrossRef](#)]
48. Fauquet, C.M.; Briddon, R.W.; Brown, J.K.; Moriones, E.; Stanley, J.; Zerbini, M.; Zhou, X. Geminivirus strain demarcation and nomenclature. *Arch. Virol.* **2008**, *153*, 783–821. [[CrossRef](#)]
49. Roumagnac, P.; Granier, M.; Bernardo, P.; Deshoux, M.; Ferdinand, R.; Galzi, S.; Fernandez, E.; Julian, C.; Abt, I.; Filloux, D.; et al. Alfalfa Leaf Curl Virus: An Aphid-Transmitted Geminivirus. *J. Virol.* **2015**, *89*, 9683–9688. [[CrossRef](#)]
50. Varsani, A.; Navas-Castillo, J.; Moriones, E.; Hernández-Zepeda, C.; Idris, A.; Brown, J.K.; Zerbini, F.M.; Martin, D.P. Establishment of three new genera in the family Geminiviridae: Becurtovirus, Eragrovirus and Turncurtovirus. *Arch. Virol.* **2014**, *159*, 2193–2203. [[CrossRef](#)]
51. Zerbini, F.M.; Briddon, R.W.; Idris, A.; Martin, D.P.; Moriones, E.; Navas-Castillo, J.; Rivera-Bustamante, R.; Roumagnac, P.; Varsani, A. ICTV Report Consortium ICTV Virus Taxonomy Profile: Geminiviridae. *J. Gen. Virol.* **2017**, *98*, 131–133. [[CrossRef](#)] [[PubMed](#)]
52. Claverie, S.; Bernardo, P.; Kraberger, S.; Hartnady, P.; Lefebvre, P.; Lett, J.-M.; Galzi, S.; Filloux, D.; Harkins, G.W.; Varsani, A.; et al. From Spatial Metagenomics to Molecular Characterization of Plant Viruses: A Geminivirus Case Study. In *Advances in Clinical Chemistry*; Elsevier: Amsterdam, The Netherlands, 2018; Volume 101, pp. 55–83.
53. Bernardo, P.; Golden, M.; Akram, M.; Naimuddin; Nadarajan, N.; Fernandez, E.; Granier, M.; Rebelo, A.G.; Peterschmitt, M.; Martin, D.P.; et al. Identification and characterisation of a highly divergent geminivirus: Evolutionary and taxonomic implications. *Virus Res.* **2013**, *177*, 35–45. [[CrossRef](#)]

54. Susi, H.; Filloux, D.; Frilander, M.J.; Roumagnac, P.; Laine, A.-L. Diverse and variable virus communities in wild plant populations revealed by metagenomic tools. *PeerJ* **2019**, *7*, e6140. [[CrossRef](#)] [[PubMed](#)]
55. Oxford Nanopore Technologies. Available online: <https://nanoporetech.com/> (accessed on 18 March 2021).
56. Porechop. Available online: <https://github.com/rrwick/Porechop> (accessed on 18 March 2021).
57. De Coster, W.; D’Hert, S.; Schultz, D.T.; Cruets, M.; Van Broeckhoven, C. NanoPack: Visualizing and processing long-read sequencing data. *Bioinformatics* **2018**, *34*, 2666–2669. [[CrossRef](#)] [[PubMed](#)]
58. Kolmogorov, M.; Yuan, J.; Lin, Y.; Pevzner, P.A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **2019**, *37*, 540–546. [[CrossRef](#)]
59. Medaka. Available online: <https://github.com/nanoporetech/medaka> (accessed on 18 March 2021).
60. Koren, S.; Walenz, B.P.; Berlin, K.; Miller, J.R.; Bergman, N.H.; Phillippy, A.M. Canu: Scalable and accurate long-read assembly via adaptivek-mer weighting and repeat separation. *Genome Res.* **2017**, *27*, 722–736. [[CrossRef](#)] [[PubMed](#)]
61. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **2018**, *34*, 3094–3100. [[CrossRef](#)] [[PubMed](#)]
62. Rozewicki, J.; Li, S.; Amada, K.M.; Standley, D.M.; Katoh, K. MAFFT-DASH: Integrated protein sequence and structural alignment. *Nucleic Acids Res.* **2019**, *47*, W5–W10. [[CrossRef](#)]
63. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree 2—Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE* **2010**, *5*, e9490. [[CrossRef](#)] [[PubMed](#)]
64. Paradis, E.; Schliep, K. ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **2018**, *35*, 526–528. [[CrossRef](#)]
65. Muhire, B.M.; Varsani, A.; Martin, D.P. SDT: A Virus Classification Tool Based on Pairwise Sequence Alignment and Identity Calculation. *PLoS ONE* **2014**, *9*, e108277. [[CrossRef](#)]
66. Yilmaz, S.; Allgaier, M.; Hugenholtz, P. Multiple displacement amplification compromises quantitative analysis of metagenomes. *Nat. Methods* **2010**, *7*, 943–944. [[CrossRef](#)]
67. Kim, K.-H.; Bae, J.-W. Amplification Methods Bias Metagenomic Libraries of Uncultured Single-Stranded and Double-Stranded DNA Viruses. *Appl. Environ. Microbiol.* **2011**, *77*, 7663–7668. [[CrossRef](#)]
68. Gallet, R.; Fabre, F.; Michalakakis, Y.; Blanc, S. The Number of Target Molecules of the Amplification Step Limits Accuracy and Sensitivity in Ultradeep-Sequencing Viral Population Studies. *J. Virol.* **2017**, *91*, 91. [[CrossRef](#)]
69. Laehnemann, D.; Borkhardt, A.; McHardy, A.C. Denoising DNA deep sequencing data—high-throughput sequencing errors and their correction. *Brief. Bioinform.* **2016**, *17*, 154–179. [[CrossRef](#)] [[PubMed](#)]
70. Wick, R.R.; Judd, L.M.; Holt, K.E. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* **2019**, *20*, 1–10. [[CrossRef](#)] [[PubMed](#)]
71. Gargis, A.S.; Cherney, B.; Conley, A.B.; McLaughlin, H.P.; Sue, D. Rapid Detection of Genetic Engineering, Structural Variation, and Antimicrobial Resistance Markers in Bacterial Biothreat Pathogens by Nanopore Sequencing. *Sci. Rep.* **2019**, *9*, 13501–13514. [[CrossRef](#)]
72. Seah, A.; Lim, M.C.; McAloose, D.; Prost, S.; Seimon, T.A. MinION-Based DNA Barcoding of Preserved and Non-Invasively Collected Wildlife Samples. *Genes* **2020**, *11*, 445. [[CrossRef](#)] [[PubMed](#)]
73. Shendure, J.; Ji, H. Next-generation DNA sequencing. *Nat. Biotechnol.* **2008**, *26*, 1135–1145. [[CrossRef](#)]
74. Ge, L.; Zhang, J.; Zhou, X.; Li, H. Genetic Structure and Population Variability of Tomato Yellow Leaf Curl China Virus. *J. Virol.* **2007**, *81*, 5902–5907. [[CrossRef](#)]
75. Harkins, G.W.; Delpont, W.; Duffy, S.; Wood, N.; Monjane, A.L.; Owor, B.E.; Donaldson, L.; Saumtally, S.; Triton, G.; Briddon, R.W.; et al. Experimental evidence indicating that mastreviruses probably did not co-diverge with their hosts. *Virol. J.* **2009**, *6*, 104. [[CrossRef](#)]
76. Domingo, E.; Sheldon, J.; Perales, C. Viral Quasispecies Evolution. *Microbiol. Mol. Biol. Rev.* **2012**, *76*, 159–216. [[CrossRef](#)]
77. Jeske, H. Barcoding of Plant Viruses with Circular Single-Stranded DNA Based on Rolling Circle Amplification. *Viruses* **2018**, *10*, 469. [[CrossRef](#)] [[PubMed](#)]
78. McNaughton, A.L.; Roberts, H.E.; Bonsall, D.; De Cesare, M.; Mokaya, J.; Lumley, S.F.; Golubchik, T.; Piazza, P.; Martin, J.B.; De Lara, C.; et al. Illumina and Nanopore methods for whole genome sequencing of hepatitis B virus (HBV). *Sci. Rep.* **2019**, *9*, 1–14. [[CrossRef](#)] [[PubMed](#)]
79. Viehweger, A.; Krautwurst, S.; Lamkiewicz, K.; Madhugiri, R.; Ziebuhr, J.; Hölzer, M.; Marz, M. Direct RNA nanopore sequencing of full-length coronavirus genomes provides novel insights into structural variants and enables modification analysis. *Genome Res.* **2019**, *29*, 1545–1554. [[CrossRef](#)] [[PubMed](#)]
80. Chang, J.J.M.; Ip, Y.C.A.; Ng, C.S.L.; Huang, D. Takeaways from Mobile DNA Barcoding with BentoLab and MinION. *Genes* **2020**, *11*, 1121. [[CrossRef](#)] [[PubMed](#)]