# Preservation of viral genomes in 700-y-old caribou feces from a subarctic ice patch

Terry Fei Fan Ng[a,b], Li-Fang Chen[c], Yanchen Zhou[a,b], Beth Shapiro[d], Mathias Stiller[d], Peter D. Heintzman[d], Arvind Varsani[e,f,g], Nikola O. Kondov[a], Walt Wong[a], Xutao Deng[a,b], Thomas D. Andrews[h], Brian J. Moorman[i], Thomas Meulendyk[j], Glen MacKay[h], Robert L. Gilbertson[c], and Eric Delwart[a,b,1]

[a]Blood Systems Research Institute, San Francisco, CA 94118; [b]Department of Laboratory Medicine, University of California, San Francisco, CA 94118; [c]Department of Plant Pathology, University of California, Davis, CA 95616; [d]Department of Ecology & Evolutionary Biology, University of California, Santa Cruz, CA 95064; [e]School of Biological Sciences and Biomolecular Interaction Centre, University of Canterbury, Christchurch, New Zealand; [f]Department of Plant Pathology and Emerging Pathogens Institute, University of Florida, Gainesville, FL 32611; [g]Electron Microscope Unit, Division of Medical Biochemistry, Department of Clinical Laboratory Sciences, University of Cape Town, Rondebosch, Cape Town, 7701, South Africa; [h]Prince of Wales Northern Heritage Centre, Government of the Northwest Territories, Yellowknife, NT, Canada X1A2L9; [i]Department of Geography, University of Calgary, Calgary, AB, Canada T2N1N4; and [j]Department of Physical and Environmental Sciences, University of Toronto, Scarborough, Toronto, Ontario, Canada M1C1A4

Edited by Peter Palese, Icahn School of Medicine at Mount Sinai, New York, NY, and approved September 30, 2014 (received for review June 6, 2014)

**Viruses preserved in ancient materials provide snapshots of past viral diversity and a means to trace viral evolution through time. Here, we use a metagenomics approach to identify filterable and nuclease-resistant nucleic acids preserved in 700-y-old caribou feces frozen in a permanent ice patch. We were able to recover and characterize two viruses in replicated experiments performed in two different laboratories: a small circular DNA viral genome (ancient caribou feces associated virus, or aCFV) and a partial RNA viral genome (Ancient Northwest Territories cripavirus, or aNCV). Phylogenetic analysis identifies aCFV as distantly related to the plant-infecting geminiviruses and the fungi-infecting *Sclerotinia sclerotiorum* hypovirulence-associated DNA virus 1 and aNCV as within the insect-infecting *Cripavirus* genus. We hypothesize that these viruses originate from plant material ingested by caribou or from flying insects and that their preservation can be attributed to protection within viral capsids maintained at cold temperatures. To investigate the tropism of aCFV, we used the geminiviral reverse genetic system and introduced a multimeric clone into the laboratory model plant *Nicotiana benthamiana*. Evidence for infectivity came from the detection of viral DNA in newly emerged leaves and the precise excision of the viral genome from the multimeric clones in inoculated leaves. Our findings indicate that viral genomes may in some circumstances be protected from degradation for centuries.**

metagenomics | reverse genetics | ancient virus | paleopathology | aDNA

Viruses infect all cellular domains of life, but descriptions of ancient autonomous viral genomes remain rare. Until recently, genetically characterized ancient RNA viruses were limited to influenza viruses from the 1918–19 outbreak, which were RT-PCR amplified from both formalin-fixed and permafrost-frozen human lung tissues and tomato mosaic tobamovirus RNA in ancient Greenland ice cores up to 140,000 y old (1–3). The genome of Barley stripe mosaic virus RNA was detected by RT-PCR in 600–900 y before present (BP) barley grains from North Africa and further sequenced using the Illumina Hi-Seq platform (4). Descriptions of ancient DNA viruses include a bacteriophage from dehydrated feces in a closed barrel buried in a 14th-century latrine in Belgium and plankton-infecting coccolithovirus phage DNA detected by PCR from sediments as old as 7,000 y under the Black sea floor (5, 6). A replication-competent pandoravirus-like Pithovirus with a 600-kb genome was reported isolated following Acantamoeba inoculation with 30,000-y-old thawed Siberian permafrost (7). Characterization of ancient viral homologs of an extant virus can increase our understanding of viral evolution (4, 5).

Alpine ice patches are unique repositories of archaeological artifacts and biological specimens. In the Selwyn Mountains, Northwest Territories, Canada, layers of caribou (*Rangifer tarandus caribou*) feces have been frozen in ice throughout the last 5,000 y. Caribou visit the ice patches during the warm summer months to obtain relief from biting insects, resulting in the accumulation of layers of fecal material (8–10). Melting ice has exposed some well-preserved biological materials from some of these ice patches. The ancient caribou feces analyzed here were located between solidly frozen ice layers more than 1 m below the surface of the ice patch (8–10). In the present study, we attempted to isolate ancient viral nucleic acids from ice patch core KfTe-1-C1-5, which contains well-preserved fecal matter and associated plant remains. Previous analyses of KfTe-1-C1-5 show that the core spans 700–4,420 radiocarbon years BP (10) (Fig. 1*A*). A previous study that attempted to PCR-amplify a long fragment of caribou mitochondrial DNA from samples from the same ice core was not successful, demonstrating that long and intact fragments of host DNA (1,306 base pairs, bp) did not survive (10). The survival of shorter fragments of mtDNA in this ice core, as would be more typical for ancient remains, has yet to be tested, however.

To evaluate the viral and host DNA content of these cryogenically preserved herbivore feces, we conducted a metagenomics analysis on the virus communities and attempted to amplify via

## Significance

**Knowledge of ancient viruses is limited due to their low concentration and poor preservation in ancient specimens. Using a viral particle-associated nucleic acid enrichment approach, we genetically characterized one complete DNA and one partial RNA viral genome from a 700-y-old fecal sample preserved in ice. Using reverse genetics, we reconstituted the DNA virus, which replicated and systemically spread in a model plant species. Under constant freezing conditions, encapsidated viral nucleic acids may therefore be preserved for centuries. Our finding indicates that cryogenically preserved materials can be repositories of ancient viral nucleic acids, which in turn allow molecular genetics to regenerate viruses to study their biology.**

Author contributions: T.F.F.N., L.-F.C., Y.Z., B.S., R.L.G., and E.D. designed research; T.F.F.N., L.-F.C., Y.Z., M.S., P.D.H., A.V., N.O.K., W.W., X.D., T.D.A., B.J.M., T.M., and G.M. performed research; T.F.F.N., T.D.A., B.J.M., T.M., and G.M. contributed new reagents/analytic tools; T.F.F.N., L.-F.C., Y.Z., B.S., M.S., P.D.H., A.V., X.D., R.L.G., and E.D. analyzed data; and T.F.F.N., L.-F.C., B.S., P.D.H., A.V., R.L.G., and E.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database [accession nos. KJ938716 (Ancient caribou feces associated virus) and KJ938718 (Ancient Northwest Territories cripavirus)].

See Commentary on page 16643.

[1]To whom correspondence should be addressed. Email: delwarte@medicine.ucsf.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1410429111/-/DCSupplemental.

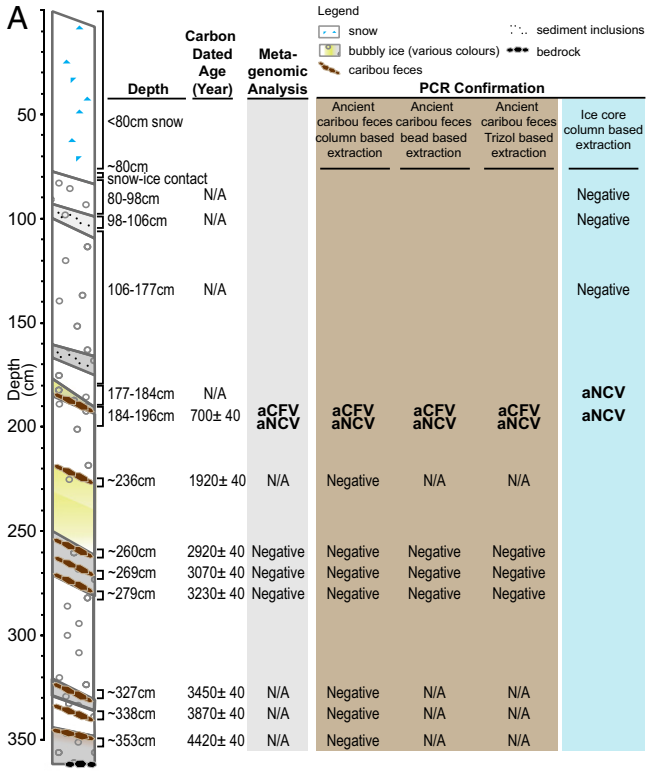


**Fig. 1.** Metagenomic identification and PCR screening of ancient viruses cryogenically preserved in frozen caribou feces and in alpine ice cores. (*A*) Schematic of KfTe-1-C1-5 ice patch core used in this study, describing the carbon dating of Caribou feces, depth of the samples, and the metagenomic and PCR result for aCFV and aNCV using three different extraction methods. The carbon dating and depth were described in a previous report (10). (*B*) Comparison of the viruses detected in the ancient caribou fecal virome with the present-day caribou fecal virome.

PCR caribou DNA from frozen fecal pellets extracted from the ice core at layers dated to 700, 2,920, 3,070, and 3,230 y BP.

## Results

We extracted nucleic acids from the filtered and nuclease-treated supernatant of resuspended frozen fecal material from a Northwest Territories ice patch core. To generate an unbiased view of the virome, we used random RT-PCR with 3′ degenerate primers to synthesize DNA (11), which we then sequenced using the Roche 454 platform, generating ∼640,000 sequence reads. We identified two sets of viral sequences from the 700-y-old sample using BLASTx similarity search against the viral RefSeq database in GenBank (Fig. 1 *A* and *B*). A DNA virus was initially identified from 11 pyrosequence reads, which we named Ancient caribou feces associated virus (aCFV). A single RNA viral sequence read was also identified with sequence similarity to a cripavirus (∼35% protein identities to *Drosophila* C virus), which was named Ancient Northwest Territories cripavirus (aNCV). No recognizable eukaryotic viral sequences were identified in the older samples (Fig. S1).

**Detection and Characterization of aCFV and aNCV.** We then recovered the genome of aCFV (GenBank accession no. KJ938716) by PCR using total nucleic acids extracted from a separate piece of the 700-y-old fecal sample as a template. Because the sequence analysis (BLASTx) indicated the likely presence of a small circular DNA virus, inverse PCR was used to obtain the genomic DNA. We used a pair of abutting PCR primers and generated a 2,231-base amplicon that was directly Sanger-sequenced by primer walking. The aCFV genome contains two ORFs, predicted to be transcribed bidirectionally (Fig. 2*A*). One of the ORFs encoded a replication-associated protein (Rep), which had the highest amino acid identities with Rep proteins of geminiviruses (26%)—a well-characterized family of plant-infecting viruses (12)—and gemycircularviruses (29%)—a recent clade of viral genomes recovered from animal feces, dragonflies, and fungi (13–15). The second ORF of aCFV shared no detectable sequence similarity with any protein in the nonredundant sequence database, but is presumed to encode the viral capsid protein (CP), as predicted by geminiviral genome organization. A long intergenic region of 232 nucleotides included a potential stem-loop structure (**CGACCTAT**CCTTAAGATTCT**ATAGGTCG**) with the underlined loop sequence presumed to be the target of the Rep activity. The aCFV loop sequence differed from the canonical nonamer sequence of geminiviruses [(T/C)A(G/C)TATTAC] or of the gemycircularvirus *Sclerotinia sclerotiorum* hypovirulence-associated DNA virus 1 (SsHADV-1, TAATATTAT).

Maximum likelihood and Bayesian inference phylogenetic analysis of the Rep protein confirmed that aCFV is distantly related to plant-infecting geminiviruses and gemycircularviruses (Fig. 2*A* and Fig. S2*A*), including a gemycircularvirus (FaGmV-13) obtained from caribou feces we collected in 2010–2011 and identified using the same viral metagenomics approach (Figs. 1*B* and 2*A*). The aCFV genome is therefore highly distinct from any sequenced present-day viruses and does not belong to any taxonomically defined group of single-stranded circular DNA viruses (16).

The RNA virus, aNCV (GenBank accession no. KJ938718), was initially identified by a single sequence read of 176 bp in length from the 700-y-old feces sample (Fig. 1). A partial genome sequence of 1,797 nt was then obtained by PCR and 3′ RACE (Fig. 2*B*). aNCV shares ∼35% amino acid sequence identity

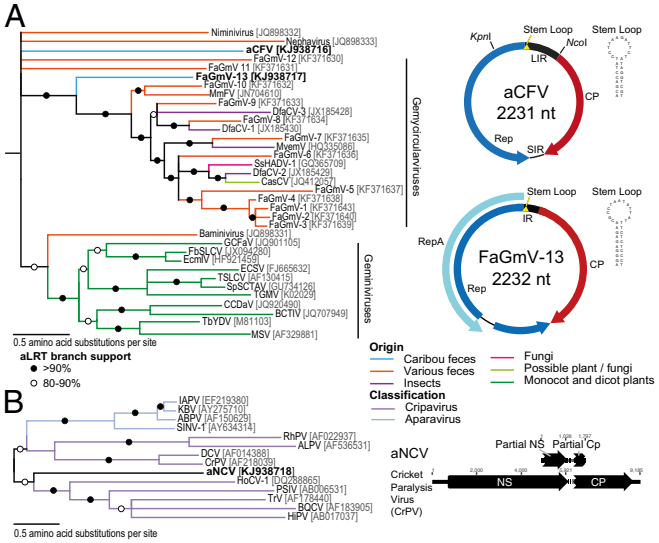


**Fig. 2.** Genome organization and phylogenetic analyses of caribou feces-derived viruses. (*A*) Phylogenetic analysis of Rep proteins in plant-infecting geminiviruses, fungi-infecting SsHADV-1, gemycircularviruses including the extant gemycircularvirus FaGmV-13 in modern caribou feces, and other circular DNA genomes including aCFV. (*B*) Phylogenetic analysis of the partial RNA-dependent RNA polymerase of aNCV with other invertebrate-infecting cripaviruses. CP, capsid protein; NS, nonstructural protein; Rep, replication-associated protein.

with the RNA dependent RNA polymerase (RdRp) protein of both *Drosophila* C virus and acute bee paralysis virus (Fig. S3). Phylogenetic analysis of the partial RdRp protein showed that aNCV belongs to the genus *Cripavirus* in the insect-infecting *Dicistroviridae* family (Fig. 2*B* and Fig. S2).

Independent experimental replication provides an important test of the authenticity of genetic analyses of ancient materials. To confirm that the ancient viruses were not contaminants introduced within the laboratory environment, we extracted and PCR-amplified nucleic acids using Qiagen viral RNA mini kits, which extract both DNA and RNA viruses (17–19), in two different laboratories using new primers and reagents (Fig. 1*A* and Fig. S4). In both laboratories, we implemented standard protocols to isolate ancient materials (20), including multiple negative controls, sterile reagents, and working in isolated, clean room environments. In all replicate experiments, the identity of PCR amplicons was confirmed by direct Sanger sequencing, whereas negative controls (no extract) consistently yielded negative PCRs. As a further control, we performed a BLASTn search of all sequence reads generated previously by deep sequencing at the Blood Systems Research Institute (BSRI). We found no prior amplification or detection of the aNCV or aCFV genomes. No previous viral amplification work was performed at the UCSC Paleogenomics Lab. These results strongly suggest that the viral genomes did not originate from the laboratory environment.

A second possible source of contamination is the silica used in the purification columns (21). We tested this possibility by extracting subsamples of the fecal samples using two additional methods: magnetic bead-based and phase-separation extraction (Fig. 1*A* and Fig. S4). These alternative methods also yielded positive PCR results for both viruses, whereas coprocessed negative controls remained negative. The older fecal samples tested negative for aCFV and aNCV using all three extraction methods and the same PCR protocols (Fig. 1*A*). A third source of potential contamination is viral particles from the present day that may be introduced either during the coring process or via passage through the ice. To test this, we processed samples of ice (no feces) from layers adjacent to the 700-y-old fecal pellets. aCFV was not detected in any ice layers tested. aNCV RNA was detected in two ice samples at ∼177–184 cm and 184–196 cm deep (Fig. 1*A*), but not in the multiple layers of ice above. Because aNCV is likely an insect virus, it is possible that this virus was deposited by flying insects on both caribou feces and adjacent snow.

To test whether host DNA was preserved within these samples, we amplified a short fragment (140 bp) of caribou mtDNA targeting the mitochondrial control region that had been amplified previously from other ancient caribou fecal pellets (22). Previously, amplification of a much longer fragment (∼1,300 bp) had been unsuccessful (10), as expected for ancient cellular DNA, which tends to degrade rapidly postmortem (23). We successfully amplified the short mtDNA fragment from the samples dating to 700 and 3,230 y BP, which includes the sample that yielded the viral sequences.

**Present-Day Caribou Feces.** We then collected contemporary caribou feces to analyze their viral content in an attempt to characterize close viral relatives of aCFV and aNCV and to measure evolutionary rates. Out of 26,150 pyrosequence reads, three viruses were identified from the six present-day caribou feces samples collected in the Northwest Territories ice patch study area in the Selwyn Mountains. The viruses, in decreasing order of reads, were sobemovirus (74 reads), tombusvirus (three reads), and gemycircularvirus (one read) (Fig. 1*B*). Sobemovirus and tombusvirus are plant-infecting RNA viruses, whereas gemycircularviruses are DNA viruses that have been characterized from animal feces, dragonflies, and fungi, but their exact cellular hosts remain unconfirmed (14). The sobemovirus sequences in present-day caribou feces shared ∼55% identity to CP sequences of previously reported sobemoviruses, the tombusvirus sequences shared ∼80% identity to tombusvirus RdRp proteins, and the gemycircularvirus sequence shared ∼60% identity to the Rep protein of other gemycircularviruses. We again used inverse PCR to amplify and Sanger-sequence the complete circular DNA genome, which we named caribou feces-associated gemycircularvirus (FaGmV-13, GenBank accession no. KJ938717). Phylogenetically, FaGmV-13 was more closely related to SsHADV-1 and gemycircularviruses than it was to aCFV (Fig. 1*A*). Like many other gemycircularviruses, its Rep was likely expressed from a spliced transcript (16), and the nonamer sequence within the stem loop was most similar to that of SsHADV-1. FaGmV-13 is a different virus species highly divergent from aCFV, therefore precluding evolutionary rate analysis. Additionally, all present-day caribou feces tested negative for aCFV using PCR.

**Inoculation of Experimental Plants with Infectious aCFV Clones.** To investigate the biology of aCFV, we used a reverse genetic approach with the laboratory plant *Nicotiana benthamiana*, a well-established model system for the study of the infectivity of cloned geminiviral DNA (24, 25). Based on the aCFV genome, we designed PCR primers overlapping an NcoI site that includes the start codon (ATG) of the putative CP ORF to amplify the full-length viral genomic DNA. PCR with this primer pair directed the amplification of the complete aCFV genome. This PCR fragment was cloned to generate the recombinant plasmids C1-monomer and C6-monomer, containing full-length clones of the aCFV genome (Fig. 3*A*). Sequence analysis of these clones revealed that the C1-monomer was identical to the aCFV genome, whereas the C6-monomer had a synonymous A486C nucleotide substitution. Multimeric viral clones (C1-1.2 mer and C6-1.2 mer) were then generated to test infectivity in plants (24).

*N. benthamiana* plants bombarded with the DNA of the multimeric clones did not exhibit any obvious symptoms (Fig. 3*F*). However, PCR of DNA extracted from newly emerged leaves of these plants 12 d postinoculation with two aCFV-specific primer pairs revealed viral DNA in ∼45% of inoculated plants (5/10 with the C1-1.2 mer and 4/10 with the C6-1.2 mer, respectively; positive result shown in Fig. S5). Negative control plants bombarded with gold particles without aCFV DNA showed no symptoms and were PCR-negative for aCFV (Fig. S5). In addition, the multimeric clones were then cloned into the binary vector pCAMBIA1300 to generate recombinant plasmids that were transformed into *Agrobacterium tumefaciens* to generate agroinoculation systems (C1-Agro and C6-Agro). *N. benthamiana* plants agroinoculated with aCFV also did not show any symptoms; however, PCR with aCFV-specific primer pairs revealed ∼90% of plants were infected (7/8 with C1-Agro and 8/8 with C6-Agro). Plants agroinoculated with the empty vector control did not develop obvious symptoms and were PCR-negative for aCFV (Fig. S5).

To further confirm virus replication in *N. benthamiana*, we performed PCR analyses to detect the replicating aCFV genome excised from the plasmid vector during replication (Fig. 3). Precise excision of the genome was confirmed in the newly emerged leaves of bombarded and agroinoculated plants (Fig. 3 *D* and *E*). Confirmation was achieved by PCR with aCFV-VB primers to direct the amplification of the expected 625-bp fragment. This result confirmed the release of viral genome from the much larger multimeric clones (Fig. 3*C*). PCR with the aCFV-VC primers directed the amplification of the 1781-bp fragment expected for the remainder of the genome. Sequencing of both PCR amplicons confirmed that the entire aCFV circular genome was released from the recombinant plasmid during infection (Fig. 3*D*).

To provide a second line of evidence for virus replication in the agroinoculated leaf tissues, restriction digestion of their rolling circle amplification (RCA) products resulted in the linearized genome of 2,231 bp, indicating release and replication of the full-length aCFV genomic DNA (Fig. S6). Together, these results further

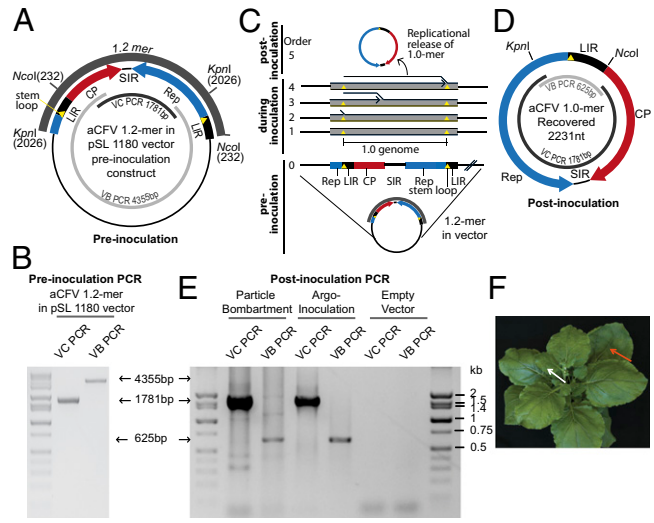


**Fig. 3.** Infectivity of the reconstituted aCFV in *N. benthamiana* plants. (*A*) Plasmid map of a 1.2-mer multimeric aCFV clone in pSL1180 vector for particle bombardment. Restriction site positions are shown relative to the original genome. (*B*) PCR result using the preinoculation plasmid, in which the outward-facing VB primers direct the amplification of a 4,355-bp fragment containing the vector. The recombinant plasmid used for agroinoculation was constructed similarly and is not shown here. (*C*) Schematic representation of the replicational release of a 1.0-mer circular genome from a 1.2-mer vector construct, adapted from the well-established geminiviral model (25). Replication is initiated at a nick in the stem loop indicated by the first yellow triangle and ended at the later stem loop at one genome length. This replicational released circular genome was then detected by PCR in Fig. 3 *D* and *E*. (*D*) Schematic representation of the recovered aCFV genomic DNA postinoculation of *N. benthamiana* plants. (*E*) PCR result of the particle bombarded and agroinoculated plants using VB and VC primers. The VB primers direct the amplification of a 625-bp fragment. The VB and VC PCR reactions demonstrated the precise excision of aCFV from the plasmid vector during replication and the circular nature of the excised genome. (*F*) *N. benthamiana* plants at 21 d postbombardment of multimeric aCFV clones. White arrow shows the location of the young newly emerged leaves taken for PCR testing; the orange arrow shows one of the leaves that was inoculated.

establish that the aCFV genomic DNA replicates in the plant species, *N. benthamiana*, via a rolling circle-type mechanism.

## Discussion

Characterizing viral genomes in ancient biological materials can help reconstruct past viral diversity and provide insights into how viruses evolve over time. Current knowledge of ancient viruses is limited due to several challenges. First, their sequences may be highly divergent and therefore difficult to identify through sequence similarities. Sequence-specific tests, such as PCR, based on current knowledge of extant viral genomes, may not detect highly divergent progenitors. Second, viral nucleic acid contents may be very low and highly degraded. In the present study, using a metagenomics and deep sequencing approach, we characterized the genomes of two divergent viruses, the DNA virus aCFV and the RNA virus aNCV, from alpine ice-preserved caribou feces radiocarbon dated to 700 ± 40 y BP. The presence of nucleic acids from both viruses was independently confirmed by two laboratories, and by different extraction methods, whereas the reagents used and extraction procedure controls were consistently PCR-negative.

Contrary to the general observation that ancient cellular DNA is highly fragmented (26), the recovered viral genomes were apparently well preserved, as reflected by the PCR amplification of the complete 2.2-kb genome of the small circular ssDNA virus (aCFV) and an 1.8-kb fragment of a ssRNA viral genome (aNCV). Both our data and a previous report (10) confirmed that the caribou cellular mtDNA within these samples was indeed highly fragmented, only allowing the amplification of a short 140-bp fragment. Comparing the long viral sequence length with the short cellular mtDNA obtained here, it reflects the preferred preservation of the viral nucleic acids. A possible explanation for the preservation of viral nucleic acids is their encapsidation within highly stable virions, thereby protecting them from degradation. Encapsidation also allows for selective enrichment using viral metagenomic methodology (17, 27, 28). The constant low temperature was likely also important in the preservation of viral nucleic acids, although local climate dynamics and the extent of freeze/thaw cycles may ultimately determine the limits of preservation, as no viral genetic material was recovered from the older samples.

Packaging nucleic acids within tight viral capsids might facilitate their long-term preservation. Both aCFV and aNCV genomes are related to viruses with icosahedral capsids. Some encapsidated viruses replicating in the gut can remain infectious in feces and other harsh environments such as sewage (17, 29). Viruses infecting animals, plants, or insects can also be consumed by animals and transit through their gastrointestinal tracts, and their nuclease-resistant, presumably encapsidated, nucleic acids are frequently reported in feces (18, 19, 30). Previous studies have also shown that enteric RNA viruses including rotavirus, hepatitis A virus, and poliovirus can remain infectious at room temperature for more than 2 mo (31). Plant viruses shed in human feces can also remain infectious (30).

Modern caribou feces contained RNA viruses related to plant-infecting tombusviruses and sobemoviruses (Fig. 1*B*), as well as a DNA virus, FaGmV-13, which is distantly related to fungi-infecting SsHADV-1 and plant-infecting geminiviruses (Fig. 2*A*). We did not recover any tombusvirus- and sobemovirus-related genomes from the 700-y-old fecal sample but instead detected a DNA virus and a member of a different RNA virus family. Because the summer plant diet of the caribou has likely remained stable over the last 5,000 y (10), the apparent difference between the ancient and modern fecal viromes may be due to differences in rate of degradation of different viral families and/or spatial and temporal variations of viruses in caribou feces collected 700 y apart.

We demonstrated that reverse genetics could be useful in studying the biology of an ancient virus (2). We investigated the infectivity of the genomic DNA of aCFV in the experimental host plant, *N. benthamiana*, by inoculation with reconstituted multimeric aCFV DNA clones. *N. benthamiana* was well suited for these studies because it is susceptible to a diverse range of plant viruses (24, 25). Although the inoculated plants became infected (Fig. 3), they did not develop obvious diseases (Fig. 3). This may be because *N. benthamiana* is a suboptimal host for replication or infection. Asymptomatic infections have been observed previously in *N. benthamiana* when inoculated with plant geminiviruses for which it is an experimental host (32–34). Here, it is also worth noting the relative simplicity of the aCFV genome—that is, only genes encoding a capsid and Rep proteins—with the symptomless phenotype reflecting the lack of gene-encoding proteins involved in movement or suppression of host defenses. To fully evaluate the pathogenicity of aCFV, if any, the original plant host will have to be identified and inoculated.

Insights into potential hosts of these ancient viruses can come from the previous paleoecological records of caribou diets and from phylogenetic analyses of extant virus-host systems. Paleoecological records indicate that the subarctic alpine ice has been a habitat for caribou for more than five millennia (8–10) and that ancient caribou populations likely migrated to ice patches for relief from heat and insect harassment during the summer months. Analyses of the fecal remains from the ice patches suggest that caribou fed on tundra vegetation in the summer, with lichens as the most abundant food source in feces (30%), followed by bryophytes

and lycopods (26.7%), shrubs (21.6%), grasses (10.5%), sedges (7.8%), and forbs (3.4%) (10). Based upon the phylogenetic analysis of its Rep gene, aCFV is distantly related to dicot- and monocot-infecting geminiviruses and the fungi-infecting SsHADV-1/gemycircularviruses (13–15). Therefore, aCFV was most likely infecting plants or fungi in this ecosystem 700 y ago. The vegetation surrounding the ice patch in the Holocene was predominantly lichens, mosses, shrubs, grasses, and sedges (10); hence, there is a high likelihood that aCFV may have infected one of these hosts. After feeding on aCFV-infected vegetation, caribous presumably defecated on the ice patches. The likely insect-infecting aNCV dicistrovirus may have also been consumed with the vegetation or originated from a caribou-harassing insect.

aCFV is highly divergent to published viral sequences resulting in a long phylogenetic branch length, possibly because no virus has yet been described from lichens, or subarctic bryophytes, monocots, and dicots. This reflects the unexplored nature of subarctic plant viruses, precluding the direct comparison of ancient genomes with their extant descendants. Future sampling in subarctic plant viromes may lead to the identification of closely related variants of aCFV; this would allow for both host identification and mutation rate measurement.

In conclusion, we demonstrate that genetic material from ancient viruses associated with caribou fecal matter was cryogenically preserved for at least seven centuries and that the cloned DNA genome of one of these viruses (aCFV) replicated and spread systemically in an extant plant. As climate change accelerates the melting of arctic ice, it is possible that ancient viral particles and the associated nucleic acids could be released into the environment. Although low temperature might preserve the genetic signature of some plant- and insect-related viruses, it is currently unknown whether such particles remained infectious. If such virions are infectious, as recently claimed for a large nucleocytoplasmic DNA virus in permafrost (7), their release could contribute to the diversity of circulating viruses.

## Methods

**Sample Collection and Dating of Ancient and Modern Caribou Feces.** Core KfTe-1-C1-5 was drilled and extracted in 2008 using a Cold Regions Research and Engineering Laboratory coring auger, which is ~7.5 cm in diameter. The extracted ice core was 330 cm in length and was described and photographed at the core site on the ice patch. Fecal samples were collected in the field using sterile tools sealed in separate sterile sample bags and frozen for later radiocarbon dating. Each unit of the rest of the ice core was individually double bagged, cooled to –15 °C, and shipped to the University of Calgary for further analysis (35). During storage and analysis, the ice core was constantly kept at or below –20 °C. Fecal samples preserved for radiocarbon dating were shipped to the Prince of Wales Northern Heritage Centre and kept frozen. Selected samples were thawed, dried in laboratory conditions, and shipped to Beta Analytic for dating (10).

Present-day samples were collected in the Northwest Territories ice patch study area in the Selwyn Mountains during March 2010 to March 2011. A total of five fecal samples were collected immediately after witnessing each caribou defecating on snow, and the sample was kept frozen from field to laboratory. Frozen caribou fecal samples collected in 2004 were also retrospectively added to the analysis. To avoid contamination, these samples were processed using viral metagenomics at least 1 y after the initial experiment on the ancient fecal samples.

**Viral Metagenomics.** We selected for metagenomic analysis four frozen caribou fecal samples from the well-studied ice core Kfte-1-C1-5 (10). The samples were dated to 700, 2,920, 3,070, and 3,230 y BP. To prevent exposure to modern contaminants, the samples were subsampled while frozen in a sterile environment at the University of Calgary and shipped, frozen, to the paleogenomics laboratory at Pennsylvania State University. There, blocks of ice containing fecal pellets were excised within the sterile, ancient DNA laboratory and stored within four layers of plastic bags and containers at –80 °C. Subsamples from each layer were shipped on dry ice to the Blood Systems Research Institute. The frozen fecal samples were unwrapped in sterilized laminar flow hood inside a high-efficiency particulate absorption filtered BSL3 room, and a small fraction (~0.2 g) was transferred to a sterile 1.5-mL centrifuge tube. Samples were each resuspended in 1 mL of sterile PBS. The resuspended material was centrifuged at 5,000 × *g* for 10 min. The supernatant was clarified by filtration using sterile 0.22-μm filters (Millipore). To reduce the background host and bacterial nucleic acids, the filtrate was treated with a mixture of nucleases consisting of 14 U of turbo DNase (Ambion), 75 U benzonase (Sigma-Aldrich), 3 U of Baseline Zero DNase (Epicentre), and 20 U of RNase A (Fermentas) in 1× DNase buffer (Ambion) at 37 °C for 90 min in a total volume of 150 μL. We extracted nucleic acids using Qiagen viral RNA mini kits (which extract both DNA and RNA viruses; Qiagen) and used these extracts to construct DNA libraries for 454 pyrosequencing by random amplification. To achieve sequence-independent amplification, we used four different primers (arbitrarily designed 20-base oligonucleotide followed by a randomized octamer sequence at the 3′ end) in separate reverse transcription (RT) reactions for each sample (11). After cDNA synthesis, we used 95 °C heat denaturation and one cycle of Klenow fragment DNA polymerase extension for second strand synthesis. This was followed by PCR amplification for 20 cycles after adding a primer (0.1 pM in a 50-μL reaction) consisting of only the 20-base fixed portion of the 3′ randomized primer. The amplification products from each primer were then size selected on an agarose gel and sequenced using the 454 GS FLX Titanium platform. The sequence of the 20-base fixed region was used as a tag to bin the resulting sequence reads to one of the four ancient core samples analyzed. Sequence analysis of the metagenomic reads was performed through a bioinformatic pipeline previously described (17). Briefly, sequences sharing more than 95% nucleotide identities over 35 bp were assembled into contigs. Raw reads, assembled contigs, and nonassembled singlets were compared with the GenBank nonredundant nucleotide and protein databases using BLASTn and BLASTx, using the E value cutoff of $10^{-4}$.

**Viral Genome Acquisition and Sequence Extension.** We generated the complete genome of the aCFV using inverse PCR with abutting primer pair aCFV-AF 5′-TTCAGCCCAATGGGTTTTAC-3′ and aCFV-AR 5′-TGATTATTGATAGTTTGATTAT-3′. We conducted touchdown PCR using LA taq (Clontech) 95 °C for 5 min, 45 cycles of [94 °C for 1 min, 58 °C minus 0.2 °C per cycle for 1 min, 72 °C for 3 min], followed by 72 °C for 10 min. To extend the aNCV RNA genome, we performed 3′ RACE amplification (Invitrogen) according to previously described protocols (17).

**Molecular Screening and Phylogenetic Analysis.** To confirm the authenticity of our results, we performed additional extraction and PCR screening. We replicated the extraction as described above at the UCSC Paleogenomics Lab and a laboratory on the UCSF Parnassus campus, where no prior amplification of these viral genomes had been performed. We tested eight different layers of ancient caribou feces and five different layers of ice. In addition, we performed extraction using two additional methodologies at BSRI: magnetic bead-based extraction (MagMAX, Ambion) and guanidinium thiocyanate–phenol–chloroform-based phase-separation extraction (TRIzol, Life Technologies). Extractions were performed according to the manufacturer's instructions. All extractions in both laboratories included negative (water) controls. To test the success of the replicate extractions, we performed aCFV-specific PCR using primers aCFV-PF 5′-AACCGTCGATCTCACAAACC-3′ and aCFV-PR 5′-GGTTTCGGCTGCAACATAAT-3′. We performed RT-PCR targeting aNCV RNA using primers aNCV-HF 5′-TTGCAGCATTGCTCACAAAT-3′ and aNCV-HR 5′-TGTAACCCATACTTGGAATAAAACAA-3′, using OneStep RT-PCR Kit (Qiagen). Nucleic acids from the six present-day caribou fecal samples were extracted using Qiagen column and tested for aCFV using primers described above. PCR conditions were as described above. Amplicons were sequenced by Sanger sequencing.

We performed maximum likelihood phylogenetic analysis using MUSCLE aligned datasets of Rep amino acid sequences of the aCFV and the partial RNA-dependent RNA polymerase of aNCV with their known homologs. We used PHYML (36) with the LG model of amino acid substitution and estimated statistical support of the resulting trees using approximate likelihood (37). Branches with <80% branch support were collapsed.

**Amplification of Caribou Mitochondrial DNA.** We tested DNA extracts of the four frozen caribou fecal samples for the presence of caribou using PCR amplification. We targeted a 139-bp fragment of the mitochondrial control region with the CB1F/R primer set of Kuhn et al. (22). PCRs were set up in the sterile facilities at the UCSC Paleogenomics Lab and comprised 1× HiFi buffer, 2.4 mM $MgSO_4$, 0.8 mg/mL rabbit serum albumin, 0.252 mM of each dNTP, 0.4 μM of each primer, 1 U Platinum HiFi Taq (Invitrogen), 2 μL of extract, and water for a final volume of 25 μL. Cycling conditions consisted of 94 °C for 12 min, 94 °C for 30 s, 55 °C for 45 s, 68 °C for 45 s, and 68 °C for 10 min, with the middle three steps repeated for a total of 50 cycles. Amplicons were cleaned using the MagNA method (38) and directly Sanger-sequenced in both directions. Sequences were processed in Lasergene 9 (DNASTAR) and identified these as a 100% match to caribou via BLAST.

**Generating Multimeric Virus Infectious Clones and Plant Inoculation.** We generated recombinant plasmids containing multimeric copies (1.2 mer) of infectious clones as described previously (24). Briefly, full-length monomeric virus clones were generated by PCR with a primer pair (aCFV-CF 5′-ACATG-**CCATGG**CGTACCGCAGGAAG-3′ and aCFV-CR 5′-GTACG**CCATGG**CATGTATG-GTCGG-3′) overlapping an NcoI site (shown in bold) at the 5′ end of the putative CP ORF. The PCR-amplified fragments were digested with Nco1 and cloned into NcoI-digested pCR-XL-TOPO vector (Invitrogen). Recombinant plasmids having the expected size insert (~2.3 kb) were identified by restriction enzyme digestion analysis and confirmed by sequencing. A 437-bp DNA fragment containing the IR was released from the monomeric clone by double digestion with NcoI and KpnI, and cloned into pSL1180 digested with NcoI and KpnI, to generate 0.2-mer clones. The full-length monomer, released by digestion with NcoI, was purified and ligated into the NcoI-digested 0.2 mer to generate a 1.2-mer construct. To test the infectivity of viral clones, DNA of the 1.2-mer construct was coated onto gold particles (39) and bombarded into leaves of *N. benthamiana* plants (four- to six-leaf stage). Gold particles alone were used as the negative control.

To generate agroinoculation systems, multimeric (1.2-mer) inserts were released from recombinant plasmids (containing the 1.2 mer) by digestion with BamHI and EcoRI and cloned into the pCAMBIA1300 binary vector. *A. tumefaciens* (strain C58) was transformed by electroporation with the resulting recombinant plasmids, and transformants were identified based on growth on antibiotic media and PCR analysis. Cell suspensions ($OD_{600}$, ~1.0) of *A. tumefaciens* strains containing binary plasmids with multimeric clones for agroinoculation experiments and a strain containing the "empty" pCAMBIA1300 vector as the negative control were used to inoculate *N. benthamiana* plants (four- to six-leaf stage) by needle puncture (40).

After bombardment or agroinoculation, plants were maintained in a locked growth chamber at 28 °C for 16 h/d and 25 °C for 8 h/d. The presence of viral DNA was determined by PCR and sequence analysis. To detect viral DNA by PCR, total genomic DNA was extracted from tissues of newly emerged leaves of plants at 21 d postinoculation using a modified method as described previously (41). To detect the Rep gene, the previously mentioned aCFV-PF and aCFV-PR primers were used; and to detect the CP gene, primers aCFV-QF 5′-ATGGCGTACCGCAGGAAGACC-3′ and aCFV-QR 5′-CGTCTGTGCCGATAACAGCGG-3′ were used. For PCR detection of the complete genome, two sets of primer pairs covering the entire genome were used: aCFV-VBF 5′-GTTTCTGATGGGGGTCCAGT-3′ and aCFV-VBR 5′-CTTTGTCTCCACGGGGATCT-3′; and aCFV-VCF 5′-GAGGCTCAGCGAGCACAT-3′ and aCFV-VCR 5′-GACGGAGGAATCCATTACCA-3′. PCR conditions were the same as the touchdown condition described above, with an extension of 1–3 min depending on product size.

A second test for virus replication was also performed. Fully expanded leaves of 6-wk-old *N. benthamiana* plants were agroinfiltrated with C1-Agro and C6-Agro (four plants were used, respectively). After 6 d, total DNA were extracted from the infiltrated leaf tissues and underwent RCA with Φ-29 DNA polymerase (TempliPhi, GE Healthcare). The resulting multimeric RCA products were digested with SmaI and SacII in separate reactions; both restriction enzymes were expected to cut the genome once and reduce the multimeric RCA product into linearized genome-size DNA fragments. The digested product was visualized in agarose gel with Ethidium Bromide.

**ACKNOWLEDGMENTS.** We thank Drs. Jay Levy and Fernando Teque at the University of California, San Francisco for use of their laboratory. We thank Erin Bickler, Christopher Markey, and Sarah Ives for proofreading. Support was provided by the Blood Systems Research Institute (T.F.F.N., Y.Z., N.O.K., W.W., X.D., and E.D.). B.S. and P.D.H. were supported by a grant from the Packard Foundation. Funding to extract the ice core came from Grant CC0018 (to T.D.A.) from Canada's commitment to the International Polar Year, plus support from the Prince of Wales Northern Heritage Centre and the University of Calgary. Field work and equipment support was provided by the Natural Sciences and Engineering Council of Canada (B.J.M.). Field work was supported by the Northern Students Training program (T.M.).

1. Taubenberger JK, Reid AH, Krafft AE, Bijwaard KE, Fanning TG (1997) Initial genetic characterization of the 1918 "Spanish" influenza virus. *Science* 275(5307):1793–1796.
2. Tumpey TM, et al. (2005) Characterization of the reconstructed 1918 Spanish influenza pandemic virus. *Science* 310(5745):77–80.
3. Castello JD, et al. (1999) Detection of tomato mosaic tobamovirus RNA in ancient glacial ice. *Polar Biol* 22(3):207–212.
4. Smith O, et al. (2014) A complete ancient RNA genome: Identification, reconstruction and evolutionary history of archaeological Barley Stripe Mosaic Virus. *Sci Rep* 4:4003.
5. Coolen MJ (2011) 7000 years of Emiliania huxleyi viruses in the Black Sea. *Science* 333(6041):451–452.
6. Appelt S, et al. (2014) Viruses in a 14th-century coprolite. *Appl Environ Microbiol* 80(9):2648–2655.
7. Legendre M, et al. (2014) Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. *Proc Natl Acad Sci USA* 111(11):4274–4279.
8. Ion PG, Kershaw PG (1989) The selection of snow patches as relief habitat by woodland caribou: *Rangifer tarandus caribou*. *Arct Alp Res* 21(2):203–211.
9. Farnell R, et al. (2004) Multidisciplinary investigations of alpine ice patches in Southwest Yukon, Canada: Paleoenvironmental and paleobiological investigations. *Arctic* 57(3):247–259.
10. Galloway JM, et al. (2012) Diet and habitat of mountain woodland caribou inferred from dung preserved in 5000-year-old alpine ice in the Selwyn Mountains, Northwest Territories, Canada. *Arctic* 65(5).
11. Victoria JG, et al. (2009) Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. *J Virol* 83(9):4642–4651.
12. King AM, Lefkowitz E, Adams MJ, Carstens EB, eds (2011) *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses* (Elsevier, San Diego, CA).
13. Rosario K, et al. (2012) Diverse circular ssDNA viruses discovered in dragonflies (Odonata: Epiprocta). *J Gen Virol* 93(Pt 12):2668–2681.
14. Sikorski A, et al. (2013) Novel myco-like DNA viruses discovered in the faecal matter of various animals. *Virus Res* 177(2):209–216.
15. Yu X, et al. (2010) A geminivirus-related DNA mycovirus that confers hypovirulence to a plant pathogenic fungus. *Proc Natl Acad Sci USA* 107(18):8387–8392.
16. Rosario K, Duffy S, Breitbart M (2012) A field guide to eukaryotic circular single-stranded DNA viruses: Insights gained from metagenomics. *Arch Virol* 157(10):1851–1871.
17. Ng TF, et al. (2012) High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage. *J Virol* 86(22):12161–12175.
18. Li L, et al. (2010) Bat guano virome: Predominance of dietary viruses from insects and plants plus novel mammalian viruses. *J Virol* 84(14):6955–6965.
19. Phan TG, et al. (2011) The fecal viral flora of wild rodents. *PLoS Pathog* 7(9):e1002218.
20. Cooper A, Poinar HN (2000) Ancient DNA: Do it right or not at all. *Science* 289(5482):1139.
21. Naccache SN, et al. (2013) The perils of pathogen discovery: Origin of a novel parvovirus-like hybrid genome traced to nucleic acid extraction spin columns. *J Virol* 87(22):11966–11977.
22. Kuhn TS, McFarlane KA, Groves P, Mooers AØ, Shapiro B (2010) Modern and ancient DNA reveal recent partial replacement of caribou in the southwest Yukon. *Mol Ecol* 19(7):1312–1323.
23. Shapiro B, Hofreiter M (2014) A paleogenomic perspective on evolution and gene function: New insights from ancient DNA. *Science* 343(6169):1236573.
24. Zhou YC, et al. (2008) Evidence of local evolution of tomato-infecting begomovirus species in West Africa: Characterization of tomato leaf curl Mali virus and tomato yellow leaf crumple virus from Mali. *Arch Virol* 153(4):693–706.
25. Stenger DC, Revington GN, Stevenson MC, Bisaro DM (1991) Replicational release of geminivirus genomes from tandemly repeated copies: Evidence for rolling-circle replication of a plant viral DNA. *Proc Natl Acad Sci USA* 88(18):8029–8033.
26. Hofreiter M, Serre D, Poinar HN, Kuch M, Pääbo S (2001) Ancient DNA. *Nat Rev Genet* 2(5):353–359.
27. Delwart EL (2007) Viral metagenomics. *Rev Med Virol* 17(2):115–131.
28. Allander T, Emerson SU, Engle RE, Purcell RH, Bukh J (2001) A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species. *Proc Natl Acad Sci USA* 98(20):11609–11614.
29. Cantalupo PG, et al. (2011) Raw sewage harbors diverse viral populations. *MBio* 2(5).
30. Zhang T, et al. (2006) RNA viral community in human feces: Prevalence of plant pathogenic viruses. *PLoS Biol* 4(1):e3.
31. Abad FX, Pintó RM, Bosch A (1994) Survival of enteric viruses on environmental fomites. *Appl Environ Microbiol* 60(10):3704–3710.
32. Cui X, Tao X, Xie Y, Fauquet CM, Zhou X (2004) A DNAbeta associated with Tomato yellow leaf curl China virus is required for symptom induction. *J Virol* 78(24):13966–13974.
33. Petty IT, Carter SC, Morra MR, Jeffrey JL, Olivey HE (2000) Bipartite geminivirus host adaptation determined cooperatively by coding and noncoding sequences of the genome. *Virology* 277(2):429–438.
34. Hagen C, et al. (2008) Biology and molecular characterization of cucurbit leaf crumple virus, an emergent cucurbit-infecting begomovirus in the Imperial Valley of California. *Plant Dis* 92(5):781–793.
35. Meulendyk T, Moorman BJ, Andrews TD, MacKay G (2012) Morphology and development of ice patches in Northwest Territories, Canada. *Arctic* 65(5):43–58.
36. Guindon S, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol* 59(3):307–321.
37. Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol* 55(4):539–552.
38. Rohland N, Reich D (2012) Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res* 22(5):939–946.
39. Gilbertson RL, et al. (1991) Cloning of the complete DNA genomes of four bean-infecting geminiviruses and determining their infectivity by electric discharge particle acceleration. *Phytopathology* 81(9):980–985.
40. Hou Y-M, Paplomatas EJ, Gilbertson RL (1998) Host adaptation and replication properties of two bipartite geminiviruses and their pseudorecombinants. *Mol Plant Microbe Interact* 11(3):208–217.
41. Dellaporta S, Wood J, Hicks J (1983) A plant DNA minipreparation: Version II. *Plant Mol Biol Rep* 1(4):19–21.