



# Amalga-like virus infecting *Antonospora locustae*, a microsporidian pathogen of grasshoppers, plus related viruses associated with other arthropods



Jesse D. Pyle<sup>a,b</sup>, Patrick J. Keeling<sup>c</sup>, Max L. Nibert<sup>a,b,\*</sup>

<sup>a</sup> Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA 02115, USA

<sup>b</sup> Harvard Ph.D. Program in Virology, Division of Medical Sciences, Harvard University, Boston, MA 02115, USA

<sup>c</sup> Department of Botany, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

## ARTICLE INFO

### Article history:

Received 21 December 2016

Received in revised form 18 February 2017

Accepted 21 February 2017

Available online 4 March 2017

### Keywords:

Amalgaviridae

Cryptic virus

Database mining

dsRNA virus

Microsporidia

Mycovirus

Ribosomal frameshifting

## ABSTRACT

A previously reported Expressed Sequence Tag (EST) library from spores of microsporidian *Antonospora locustae* includes a number of clones with sequence similarities to plant amalgaviruses. Reexamining the sequence accessions from that library, we found additional such clones, contributing to a 3247-nt contig that approximates the length of an amalga-like virus genome. Using *A. locustae* spores stored from that previous study, and new ones obtained from the same source, we newly visualized the putative dsRNA genome of this virus and obtained amplicons yielding a 3387-nt complete genome sequence. Phylogenetic analyses suggested it as prototype strain of a new genus in family Amalgaviridae. The genome contains two partially overlapping long ORFs, with downstream ORF2 in the +1 frame relative to ORF1 and a proposed motif for +1 ribosomal frameshifting in the region of overlap. Subsequent database searches using the predicted fusion protein sequence of this new amalga-like virus identified related sequences in the transcriptome of a basal hexapod, the springtail species *Tetrodontophora bielanensis*. We speculate that this second new amalga-like virus (contig length, 3475 nt) likely also derived from a microsporidian, or related organism, which was associated with the springtail specimens at the time of sampling for transcriptome analysis. Other findings of interest include evidence that the ORF1 translation products of these two new amalga-like viruses contain a central region of predicted  $\alpha$ -helical coiled coil, as recently reported for plant amalgaviruses, and transcriptome-based evidence for another new amalga-like virus in the transcriptome of another basal hexapod, the two-pronged bristletail species *Campodea augens*.

© 2017 Elsevier B.V. All rights reserved.

## 1. Introduction

*Antonospora locustae* is a single-celled, spore-forming obligate intracellular parasite and member of the diverse phylum Microsporidia (Slamovits et al., 2004). Microsporidia infect a broad range of animal hosts and are now generally considered to constitute a basal clade either within or closely allied with kingdom Fungi (Stentiford et al., 2016). The reduced and compacted genomes of microsporidia (as small as 2.3 Mbp; *A. locustae*, 5.4 Mbp) have garnered substantial interest with regard to their evolution (Keeling and Slamovits, 2004). At least 15 microsporidian species, though not *A. locustae*, have been identified as human pathogens, almost always in people who are severely immunocompro-

mised (<http://www.cdc.gov/dpdx/microsporidiosis/>). *A. locustae* is a pathogen and biocontrol agent for grasshoppers and other orthopterans (Lange, 2005), representing the active component of commercial products Nolo Bait and Semaspore.

An EST library prepared from purified spores of *A. locustae* ATCC30860 has been reported by Williams et al. (2005), including 1146 clones for which sequences have been deposited in the EST sequence database at GenBank/EMBL-EBI/DDBJ (accession nos. DQ057484-DQ057579 and DQ071178-DQ071262). At that time, 275 of these accessions (24%) were found to have no discernible homologs in database searches. More recently, Liu et al. (2012) have reported that a number of these accessions score as hits in EST database searches when RdRp sequences from plant amalgaviruses (monosegmented, putative dsRNA viruses that constitute genus *Amalgavirus* in family Amalgaviridae (Liu and Chen, 2009; Martin et al., 2011; Sabanadzovic et al., 2009, 2010)) are used as queries. Liu et al. (2012) have also reported that a contiguous sequence assembly (contig) of 2377 nt can be generated from 98 of these accessions

\* Corresponding author at: Department of Microbiology and Immunobiology, Harvard Medical School, 77 Ave. Louis Pasteur, Boston, MA 02115, USA.

E-mail address: [mnibert@hms.harvard.edu](mailto:mnibert@hms.harvard.edu) (M.L. Nibert).

and appears to encode an RdRp phylogenetically related to those of plant amalgaviruses.

Also related to plant amalgaviruses is a monosegmented, putative dsRNA virus from the ascomycetous yeast *Zygosaccharomyces bailii*, *Zygosaccharomyces bailii* virus Z (ZbV-Z) (Schmitt and Neuhausen, 1994), which has not been formally classified to date. Recently, we have redetermined and amended the complete genome sequence of ZbV-Z from *Z. bailii* 412, providing further support for its relationship to plant amalgaviruses (Depierreux et al., 2016). We have additionally proposed ZbV-Z/412 as prototype strain of a new genus, “Zybavirus”, in family *Amalgaviridae*. There is thus precedent for expanding *Amalgaviridae* to include viruses that infect fungi, as well as ones that infect plants.

Plant amalgaviruses and ZbV-Z are examples of RNA viruses that persistently infect their respective hosts and appear to cause limited harm, leading these and other such viruses to be labeled as “cryptic”. Cryptic viruses often lack the capacity for efficient horizontal transmission via extracellularly released or vector-borne virions. They are instead transmitted vertically or horizontally through means involving intracellular virions or related forms, such as during cell division or following cell–cell fusion events related to sexual reproduction, sometimes with subsequent segregation into spores or seeds. Moreover, many cryptic viruses seem to have evolved toward constrained levels of replication, such that host cells are not overwhelmed by viral demands on their resources or other harmful effects. Notably also, many cryptic viruses of fungi or protozoa can either increase or decrease the virulence of their hosts for their own respective animal or plant hyperhosts, suggesting that a complex web of relative harms and benefits has likely influenced the evolution of these multi-level symbioses. The term “hyperhost” has been previously suggested for denoting the host of a host (Issi, 2002), in parallel with the term “hyperparasite” for denoting the parasite of a parasite (Morozov et al., 2007).

Given the previous EST-based findings for an amalga-like virus from *A. locustae*, we decided to investigate further, in an effort to establish the existence and complete genome sequence of this virus and to refine its taxonomic placement. To our knowledge, no other viruses of microsporidia have been described to date, which might not be surprising given that large-scale sequencing efforts have so far mostly targeted the DNA genomes of microsporidia and would not have detected most types of RNA viruses. Cryptic RNA viruses might thus be more widely distributed among microsporidia, and possibly other basal fungi, than currently known.

## 2. Materials and methods

### 2.1. Database searching and contig assembly

All sequence database searches were performed with the indicated programs at <http://blast.ncbi.nlm.nih.gov/Blast.cgi>. Searches of the EST or the Transcriptome Shotgun Assembly (TSA) database with protein or nucleotide queries were performed using tblastn, megablast, or discontinuous megablast. Searches of the Non-redundant Protein Sequences (NR) database with protein queries were performed using PSI-BLAST (Schäffer et al., 2001). Searches of the Nucleotide Collection (NR/NT) or the Sequence Read Archive (SRA) database with nucleotide queries were performed using discontinuous megablast or blastn.

Analyses yielding the AnloV1 contig began with a tblastn search of the EST database for *A. locustae* (NCBI taxonomic identifier 278021), using the predicted ORF1+2p sequence of blueberry latent virus (BLV) (GenBank HM029246) as query. This search identified 46 accessions with E-values  $\leq 4e-7$ , which were assembled into a single, 1738-nt contig by CAP3 (Huang and Madan, 1999) as implemented with defaults at <http://mobyli.pasteur.fr/>. This contig was

next used as query for a megablast search of the EST database for *A. locustae*, yielding 80 accessions with E-values  $\leq 2e-33$ , which were assembled into three contigs by CAP3. Each of these three contigs was noted to have a poly(A) or poly(T) sequence near one end, followed or preceded, respectively, by a region of the multiple cloning sequence from pcDNA3.1, the vector used by Williams et al. (2005) for EST cloning. We therefore trimmed all terminal and periterminal poly(A/T) and multiple cloning sequences from the accessions that contained them and repeated the CAP3 assembly, this time obtaining a single, 2324-nt contig. The steps yielding the 2324-nt contig from the 1738-nt contig were next reiterated, starting with the 2324-nt contig as query. Each of these iterations increased the length of the contig until a 3169-nt contig (assembled from 134 accessions with E-values  $\leq 3e-33$ ) was reached as the limit product. We next examined two unincorporated accessions that had recurrently appeared in the preceding iterations. A portion of one of these accessions mapped internally to the 3169-nt contig, but the other of these accessions overlapped one end of the 3169-nt contig by 671 nt of identical sequence and then extended the contig by 78 nt of new sequence, to 3247 nt in overall length. This extended contig was next used as query for a discontinuous megablast search of the EST database for *A. locustae*, again yielding 134 hits and implying no further extension. CAP3 standard output was used to evaluate coverage and variation at each position in the final contig.

Analyses yielding the partial TebiAV1 and CaauAV1 sequences are mostly described in Results. Briefly, a tblastn search of the TSA database for arthropods (NCBI taxonomic identifier 6656), using the predicted AnloAV1 ORF1+2p sequence as query, identified two high-scoring accessions from *Tetradontophora bielanensis* (GAXI02024117 and GAXI02024802; respective E-values,  $8e-72$  and  $6e-52$ ), corresponding to the two halves of AnloAV1. The next highest-scoring accession in this tblastn search, below the two from *T. bielanensis*, was from the two-pronged bristletail *Campodea augens* (GAYN02026491; E-value,  $5e-23$ ). Additional searches of the TSA and SRA databases identified no other sequences that could be used to extend either the TebiAV1 or the CaauAV1 contig. CAP3 standard output or CLC Genomics Workbench 8.0 was used to evaluate coverage at each position in these contigs (Table S1).

### 2.2. *A. locustae* spores and RNA purifications

Purified spores of *A. locustae* isolate ATCC30860 were purchased from M&R Durango, Inc. Total RNAs were purified from these spores using previously described methods (Depierreux et al., 2016), though with some modifications. Approximately  $1e8$  spores were subjected to vortexing with 0.25 g glass beads and 8 U zymolase (Zymo Research) at room temperature for 1 h. Following disruption of the spore cell walls, samples were mixed with 500  $\mu$ L TRIzol reagent (Ambion) and 100  $\mu$ L chloroform. Samples were subjected to 15 s additional vortexing, followed by centrifugation at 13,000g for 30 min at 4°C. The aqueous phase was then removed, and total RNA was precipitated with 1 mL cold 100% isopropanol and incubation on dry ice for 10 min. The precipitated RNA was pelleted by centrifugation at 13,000  $\times$  g for 20 min at 4°C, washed once with 70% ethanol, and allowed to air-dry for 5 min at room temperature, after which this dried pellet was resuspended in 100  $\mu$ L DEPC-treated water. The purified RNA was quantified by UV spectrophotometer analysis (Quawell) and its quality assessed by agarose gel electrophoresis and ethidium bromide staining.

Total dsRNAs were enriched from total RNA samples using the microcrystalline cellulose method (Castillo et al., 2011), largely as described by Depierreux et al. (2016). To confirm the chemical nature of the enriched dsRNAs, DNase I and RNase A treatments were used. Approximately 500 ng dsRNA from *A. locustae* spores was incubated with 2 U DNase I (NEB) and 1X DNase I reaction buffer for 10 min at 37°C per manufacturer's instructions, after

which the reaction was inactivated at 75 °C for 10 min. RNase A treatments were performed as described previously (DePaulo and Powell, 1995), with minor modifications. Approximately 500 ng dsRNA from *A. locustae* spores was incubated with RNase A to a final concentration of 10 ng/mL under “high salt” (0.3 M NaCl) or “low salt” (no NaCl) conditions for 10 min at room temperature. In each case, after the DNase I or RNase A treatments, the total remaining dsRNAs were analyzed by 1% agarose gel electrophoresis and ethidium bromide staining.

### 2.3. De novo Sanger sequencing

Approximately 500 ng enriched dsRNA from *A. locustae* were used for first-strand cDNA synthesis, RT-PCR, and sequencing of AnloV1-derived amplicons. First-strand synthesis was performed using SuperScript III reverse transcriptase (Thermo Fisher Scientific) and random hexamer primers per manufacturer’s instructions except for use of an elevated initial denaturing temperature (95 °C) to aid in melting the dsRNA for primer binding. Resulting cDNAs were then used as templates for PCR with various AnloV1-specific primer pairs (Table S2) designed from the 3247-nt EST contig. Each PCR was performed using Taq polymerase (NEB). All amplicons were visualized by 1% agarose gel electrophoresis and ethidium bromide staining, after which they were excised and purified using QIAquick Gel Extraction Kits (Qiagen) per manufacturer’s instructions. Once purified, the amplicons were sent for high-throughput Sanger sequencing at the Dana Farber/Harvard Cancer Center DNA Resource Core facility, using the appropriate forward and reverse primers for sequencing from both strands of each amplicon.

RNA ligase-mediated rapid amplification of cDNA ends (RLM-RACE) was performed to determine the terminal sequences of the AnloV1 genome using previously described methods (Coutts and Livieratos, 2003; Depierreux et al., 2016). Following ligation of the DNA adapter (P-adapter-N, Table S2) to denatured dsRNA, reverse transcription was performed in parallel using both random hexamers and adapter-specific primers for first-strand synthesis. Internal AnloV1-specific (AnloV1\_R1 and AnloV1\_F5) and adapter-specific (anti-adapter-outer) primers (Table S2) were used for a first round of PCR, followed by a second round of PCR with AnloV1-specific (AnloV1\_RACE.R.inner, AnloV1\_RACE.F.inner) and adapter-specific (anti-adapter-inner) nested primers (Table S2). The resulting amplicons were analyzed and sequenced as described for other amplicons above. In the end, the complete genome sequence of AnloV1 was confirmed by bidirectional sequencing reads except for 35 nt at the plus-strand 5′ end and 25 nt at 3′ end, which were determined from outwardly directed reads only.

### 2.4. Sequence-based analyses

ORFs were identified in nucleotide sequences using EMBOSS getorf as implemented with defaults at <http://www.bioinformatics.nl/emboss-explorer/>. Molecular weight and pI values for proteins were calculated using Compute pI/MW as implemented with defaults at [http://web.expasy.org/compute\\_pi/](http://web.expasy.org/compute_pi/). Multiple sequence alignments of protein sequences were performed using PROMALS3D (Pei et al., 2008) as implemented with defaults at <http://prodata.swmed.edu/>, MAFFT-L-INS-i 7.27 (Katoh and Standley, 2013) as implemented with defaults at <http://mafft.cbrc.jp/alignment/server/>, or MUSCLE 3.8 (Edgar, 2004) as implemented with defaults at <http://www.ebi.ac.uk/Tools/msa/>. Global pairwise alignments of protein sequences were performed using Needle or NeedleX as implemented with defaults at <http://www.bioinformatics.nl/emboss-explorer/>. The ORF2p (RdRp) sequences used for multiple sequence alignments or global pairwise alignments began with the first residue after

the site of predicted PRF in ORF2 for plant amalgaviruses, AnloV1, TebiAV1, ZbV-Z, UvNV1-like viruses, and unirenaviruses; with the first in-frame Met in the RdRp-encoding ORF for CTTV-like viruses; and with the first residue in the partial RdRp-encoding ORF for CaauAV1. The ORF1p sequences used for global pairwise alignments began with the first in-frame Met in ORF1 for plant amalgaviruses, AnloV1, and ZbV-Z, and with the first residue in ORF1 for TebiAV1 (because its sequence is thought to be N-terminally truncated). Local pairwise alignments were performed using Matcher as implemented with defaults at <http://www.bioinformatics.nl/emboss-explorer/>. Coiled coil predictions were obtained using Marcoil (Delorenzi and Speed, 2002) as implemented with defaults at <https://toolkit.tuebingen.mpg.de/marcoil> and Paircoil2 (McDonnell et al., 2006) as implemented with defaults at <http://groups.csail.mit.edu/cb/paircoil2/paircoil2.html>. RNA secondary structures were predicted using the RNAfold server (Gruber et al., 2008) at <http://rna.tbi.univie.ac.at/> (with default options except temperature = 25 °C). The predicted structure was displayed using VARNAV3-93 (Ponty and Leclerc, 2015).

Phylogenetic relationships were determined using PhyML 3.0 (Guindon et al., 2010) as implemented at <http://www.hiv.lanl.gov/content/sequence/PHYML/interface.html> with the following parameters differing from the defaults: Sequence type/model, Amino acids/JTT, LG, rtREV, or WAG; Proportion of invariable sites, estimated from data; Gamma shape parameter, estimated from data; Starting tree(s) optimization, Tree topology and Branch length; Tree improvement, Best of NNI and SPR; Branch support, Approximate Likelihood Ratio Test (aLRT), SH-like supports. The output in Newick format was then opened in FigTree v1.4.0 (downloaded from <http://tree.bio.ed.ac.uk/software/figtree/>) for refining the phylogram for presentation. For Fig. 3A (MUSCLE, rtREV), values estimated from the data were Proportion of invariable sites, 0.013, and Gamma shape parameter, 2.129; for Fig. S1 (MAFFT, rtREV), values estimated from the data were Proportion of invariable sites, 0.013, and Gamma shape parameter, 1.974; and for Fig. S2A (MAFFT, rtREV), values estimated from the data were Proportion of invariable sites, 0.014, and Gamma shape parameter, 2.209. For each of these figures, alternative use of the JTT, LG, or WAG substitution model yielded results very similar to those shown. Tables 1, S1, and S3 list names, abbreviations, and GenBank accession numbers for the nucleotide sequences of all dsRNA viruses included in this study.

## 3. Results

### 3.1. Amalga-like virus sequences among EST clones from *A. locustae*

To begin this study, we searched the *A. locustae* EST database for amalgavirus-related accessions beyond the 98 described in Liu et al. (2012). We found 36 more such accessions, for 134 total, all derived from the EST library of *A. locustae* ATCC30860 (Williams et al., 2005). The 134 accessions were then assembled into a 3247-nt contig, intermediate in length to the genomes of plant amalgaviruses (3427–3437 nt) (Liu and Chen, 2009; Martin et al., 2011; Sabanadzovic et al., 2009, 2010) and amalga-like mycovirus ZbV-Z (3160 bp) (Depierreux et al., 2016). Only 114 positions in this contig were represented in fewer than 5 accessions: 8 positions at one end and 106 positions at the other. At other positions, coverage depths ranged from 5 to 52. When we then used this contig to query the genome sequence of *A. locustae* HM2013 (<http://genome.jgi.doe.gov/Antlo1/Antlo1.home.html>), we obtained no significant hits (E-values  $\geq 7e-1$ ), suggesting that it represents an element from outside the *A. locustae* genome,

likely an amalga-like virus per se. We henceforth refer to this element as *Antonospora locustae* virus 1 (AnloV1) and specifically as AnloV1 strain ATCC30860 (AnloV1/ATCC30860). The 3247-nt partial genome sequence for AnloV1/ATCC30860 is provided in supplementary file Data S1.

### 3.2. Visualization of AnloV1 dsRNA and de novo Sanger sequencing

Purified spores of *A. locustae* ATCC30860 were obtained from the same batch, continuously stored at  $-80^{\circ}\text{C}$ , as used for generating the EST library (Williams et al., 2005). Total RNA was prepared by TRIzol extraction and used as template for RT-PCR with primers designed from the 3247-nt EST contig. Results of these efforts were amplicons that yielded 3174 nt of sequences identical to those of the EST contig, representing complete sequences derived from both strands of the AnloV1/ATCC30860 genome except for small regions at the two ends as identified below.

The spore-derived total RNA was next enriched for dsRNA by cellulose affinity. When analyzed by agarose gel, two faint bands in the 2–4 kbp range appeared to be seen. For obtaining larger amounts of RNA, additional *A. locustae* ATCC30860 spores were obtained from the same commercial source (M&R Durango, Inc.) as for the Williams et al. (2005) study. Following total RNA extraction, dsRNA enrichment by cellulose affinity, and analysis by agarose gel, a band of the appropriate  $M_r$  (between 3 and 4 kbp) was readily seen, as well as a band of lower  $M_r$  (~2 kbp) (Fig. 1A). Neither band was degraded by DNase I or by RNase A in higher-salt conditions, but both were degraded by RNase A in lower-salt conditions (Fig. 1B), consistent with their identifications as dsRNA molecules. The nature of the lower- $M_r$  (~2-kbp) dsRNA band remains unproven, but preliminary RNAseq (Illumina) results from the enriched dsRNA sample have not only confirmed the AnloV1 sequence but also yielded a highly represented ~2000-nt contig that appears to represent an RNA satellite, with limited protein-coding capacity and no discernible sequence similarity to AnloV1/ATCC30860, ribosomal RNAs, or other *A. locustae* HM2013 genes. Further experiments to confirm these results and interpretations are pending.

We next used the spore-derived dsRNA to perform RLM-3'RACE for determining sequences at the AnloV1/ATCC30860 genome termini. The results revealed 140 nt beyond the ends of the 3247-nt EST contig: 11 nt at the plus-strand 5' end and 129 nt at the plus-strand 3' end. Internal sequences read from the RACE amplicons were identical to those of the EST contig. The overall length of the genomic plus strand is thus 3387 nt, approaching the genome lengths of plant amalgaviruses. This newly determined, complete

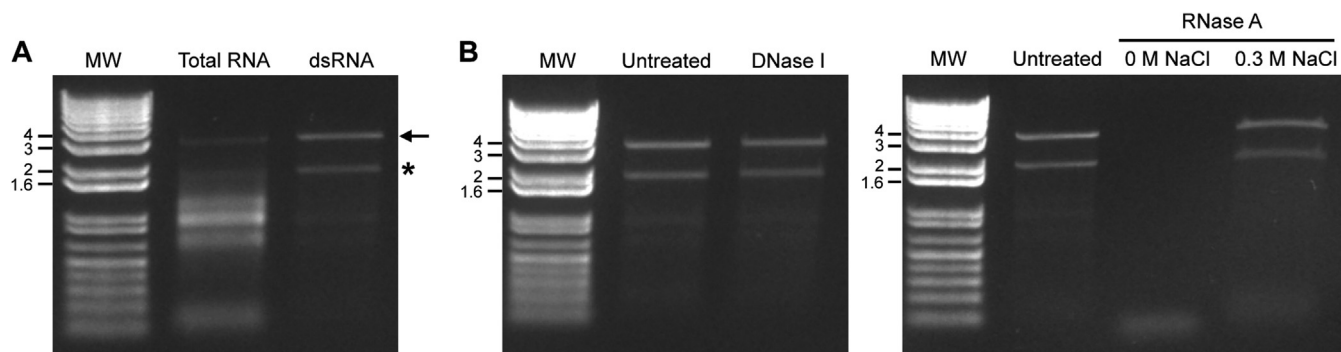
genome sequence for AnloV1/ATCC30860 has been deposited in GenBank with accession number KX525322.

### 3.3. Primary features of AnloV1 genome sequence

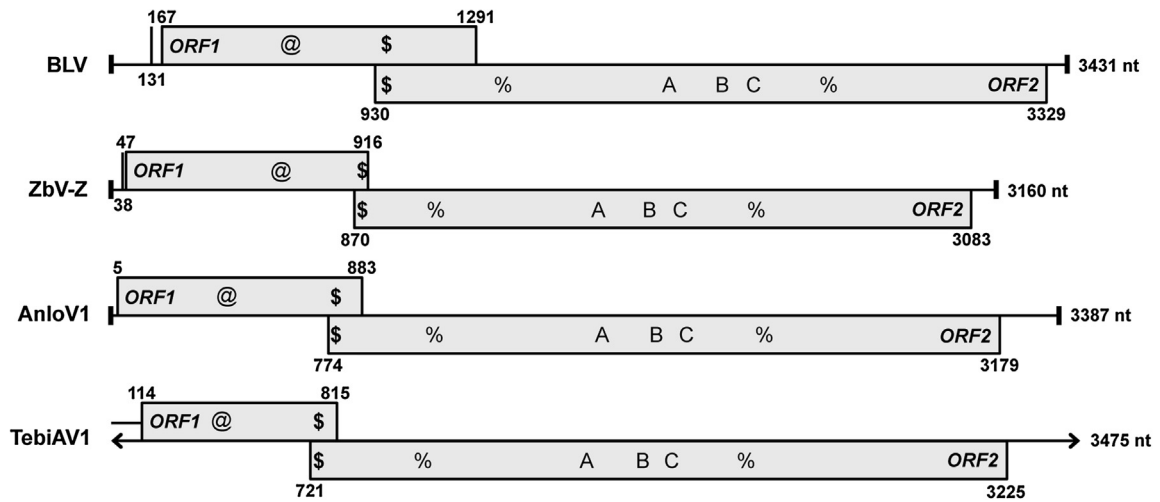
The minus strand of AnloV1/ATCC30860 is devoid of long ORFs, including none  $\geq 285$  nt in length between flanking stop codons. The plus strand, on the other hand, contains two long ORFs, which are partially overlapping (Fig. 2). ORF1 spans positions <1–883, extending from the 5' end to a flanking UGA stop codon at positions 884–886 (Table 1). This ORF1 sequence is notable for having four in-frame AUG codons within its 5'-most 109 nt: AUG1–4 at positions 5–7, 35–37, 80–82, and 107–109, followed by AUG5 at positions 224–226. Which of these AUG codons represents the functional start codon for ORF1p (translation product of ORF1) remains to be determined. For purposes here, however, we have adopted the common convention of identifying the first in-frame AUG codon (AUG1) as the predicted start codon. Thus, assuming that the protein-coding region of ORF1 spans positions 5–883, it is predicted to encode a 293-aa, 34-kDa product (pI, 8.5) (Table 1). ORF2 spans positions 774–3179, extending between a UGA stop codon at positions 771–773 and a UAA stop codon at positions 3180–3182 (Table 1). ORF1 and ORF2 thus overlap by 110 nt, not including the flanking stop codons. Notably, ORF2 is in the +1 frame relative to ORF1, as is also the case for plant amalgaviruses (Liu and Chen, 2009; Martin et al., 2011; Sabanadzovic et al., 2009, 2010) and ZbV-Z (Depierreux et al., 2016). Moreover, within the ORF1–ORF2 region is the sequence UUU.CUU.G (underlines, ORF1 codon boundaries) at positions 776–782, representing a probable +1 slippery sequence per Firth et al. (2012) and Depierreux et al. (2016) (also see Discussion). If +1 PRF indeed occurs in this motif (translation of slippage codon UUU/UUC, then UUG), then the resulting ORF1+ORF2 fusion is predicted to span positions 5–778:780–3179 and to encode a 1058-aa, 122-kDa product (pI, 9.0) that we designate ORF1+2p (Table 1). The lengths of the predicted ORF1p and ORF1+2p products of AnloV1/ATCC30860 are thus similar to those of plant amalgaviruses and ZbV-Z (Table 1).

### 3.4. Comparisons and phylogenetic analyses of AnloV1 genome sequence

In searches against the full NR database, the predicted ORF1p sequence of AnloV1/ATCC30860 yielded no significant hits (all E-values  $\geq 0.52$ ). The predicted ORF1+2p sequence, on the other hand, yielded significant hits with viral RdRp sequences, those of plant amalgaviruses and ZbV-Z constituting the top 11 hits (E-values,  $1\text{e}-38$  to  $6\text{e}-18$ ). To begin to address the phylogeny of AnloV1,



**Fig. 1.** Isolation and characterization of dsRNA from *A. locustae* spores. MW, lanes containing a set of DNA molecular weight markers (labeled in kbp). (A) Total RNA was extracted from *A. locustae* spores, and dsRNA was then enriched by cellulose affinity (see Materials and Methods). Samples of total RNA and enriched dsRNA were then separated by electrophoresis on a 1% agarose gel and visualized by ethidium bromide staining. Arrow: dsRNA band attributed to AnloV1 migrating near 3.5 kbp. Asterisk: additional dsRNA band of unknown identity (suspected satellite RNA) migrating near 2 kbp. (B) Samples of enriched dsRNA from *A. locustae* spores were subjected to treatment with DNase I (left), or RNase A (right) at either low (0 M) or high (0.3 M) NaCl concentrations.



**Fig. 2.** Amalgaviruses genome diagrams. In each genome, ORF2 is in the +1 frame relative to ORF1. Terminal bars or arrows indicate whether terminal sequences respectively have or have not been determined by RACE. Position numbers indicate the beginning and end of each ORF: from the first in-frame AUG codon to the flanking stop codon for ORF1, between the two flanking stop codons for ORF2. Position number is also shown at bottom for the beginning of ORF1 as defined by its upstream flanking stop codon, if present (none in AnloV1 and TebiV1). The first in-frame AUG codon in ORF1 is considered to be the functional start codon in BLV, ZbV-Z, and AnloV1, but because of the shorter ORF1p that would result from use of this codon in TebiAV1, the functional start codon of TebiAV1 is thought to remain upstream of the sequence determined to date. The position number of the end of ORF2 in ZbV-Z was misstated in Depierreux et al. (2016), but the number shown here and in GenBank KU200450 is the correct one. Symbols (see text for more explanations): \$, +1 PRF motif in the ORF1–ORF2 overlap region; @, central region of coiled coil propensity in ORF1p; %, two proposed amalgaviruses signature motifs in the RdRp region of ORF1+2p; ABC, core polymerase motifs in the RdRp region of ORF1+2p.

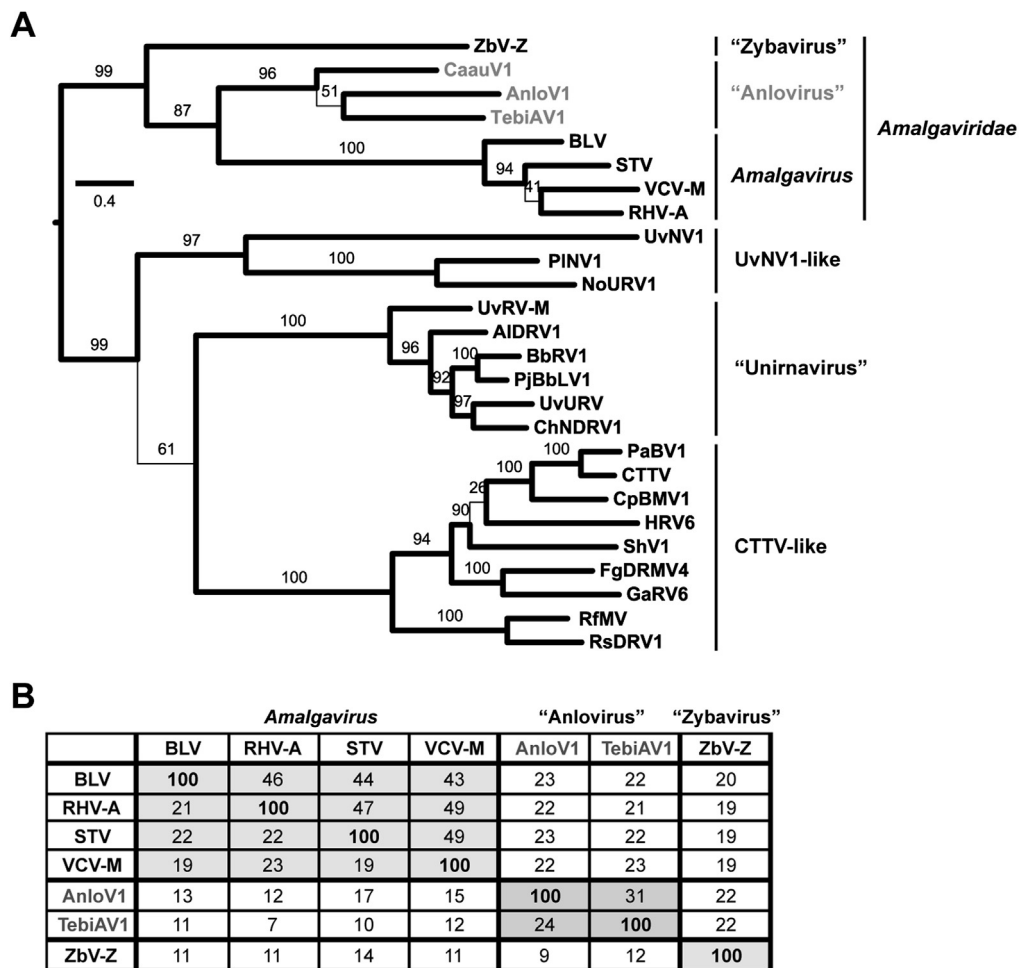
**Table 1**  
Genome-derived properties of plant amalgaviruses, ZbV-Z, and new amalgaviruses described in this report.

Virus abbrev. <sup>a</sup>	Accession no.	Genome (bp)	ORF1 range <sup>b</sup>	ORF2 range <sup>b</sup>	FS <sup>c</sup>	ORF1+ORF2 range <sup>d</sup>	NTRs <sup>e</sup>		ORF1p (aa)	ORF1+2p (aa) <sup>f</sup>
							5′	3′		
BLV	HM029246 <sup>g</sup>	3431	131–1291	930–3329	+1	167–961:963–3329	166	102	375	1054
RHV-A	HQ128706	3427	38–1306	696–3326	+1	95–867:869–3326	94	101	404	1077
STV	EF442780 <sup>g</sup>	3437	87–1268	976–3324	+1	138–1001:1003–3324	137	113	377	1062
VCV- M	EU371896	3434	20–1324	966–3314	+1	143–1000:1002–3314	142	120	394	1057
ZbV- Z	KU200450	3160	38–916	870–3083	+1	47–910:912–3083	46	77	290	1012
AnloV1	KX525322 <sup>h</sup>	3387	<1–883	774–3179	+1	5–778:780–3179	4	208	293	1058
TebiAV1	GAXI02024117 <sup>j</sup> (3475) <sup>k</sup> GAXI02024802 <sup>j</sup>	(3475) <sup>k</sup>	(<1)–815	721–3225	+1	(<1)–731:733–3325	(0)	(250)	(271)	(1074)
CaauAV1	GAYN02026491 <sup>j</sup> (440) <sup>k</sup>	(440) <sup>k</sup>	na	(1–440)	na	na	na	na	na	na

<sup>a</sup> RHV-A, rhododendron virus A; STV, southern tomato virus; VCV-M, Vicia cryptic virus M; see text for other abbreviations.  
<sup>b</sup> In these columns, the nt position range of each ORF reflects the span between flanking stop codons, with no effort to predict which start codon may be used. The flanking stop codons are not included in the specified range.  
<sup>c</sup> FS, frame shift: the frame in which ORF2 is found relative to ORF1 (either –1, 0, or +1).  
<sup>d</sup> In this column, the ORF1 5′ end is defined by the predicted start codon. The colon represents position of the predicted +1 PRF event.  
<sup>e</sup> NTRs, non-translated regions terminal to ORF1 (5′) and ORF2 (3′), including the ORF2 stop codon.  
<sup>f</sup> Based on ORF1 + ORF2 range defined in a preceding column.  
<sup>g</sup> Sequences of representative strains were used for these viruses.  
<sup>h</sup> This new accession number is for the complete genome sequence of AnloV1 determined by de novo sequencing in this study.  
<sup>j</sup> These accession numbers are for the original TSA database hits for these putative new viruses. The lengths and other values shown here, however, are based on the reassembled sequences for each as described in the text and summarized in Table S1.  
<sup>k</sup> The TebiAV1 sequence appears to have been incompletely sequenced at one or both ends, as indicated by the parentheses. The CaauAV1 sequence is clearly incomplete at both ends; na, not available.

we compared the ORF2 (RdRp) portion of its predicted ORF1+2p sequence with those of a larger group of viruses whose RdRp (and potentially CP regions as well (Nibert et al., 2016)) are related to those of plant amalgaviruses, namely, (i) members of the four approved species of plant amalgaviruses (genus *Amalgavirus*, family *Amalgaviridae*; species *Blueberry latent virus*, *Rhododendron virus A*, *Southern tomato virus*, and *Vicia cryptic virus M*) (Liu and Chen, 2009; Martin et al., 2011; Sabanadzovic et al., 2009, 2010), (ii) ZbV-Z (proposed genus “Zybavirus”, family *Amalgaviridae*) (Depierreux et al., 2016), (iii) three members of an unclassified taxon of monosegmented, putative dsRNA viruses represented by *Ustilaginoidea virens* nonsegmented virus 1 (also *Nigrospora oryzae* nonsegmented RNA virus 1 and *Purpureocillium lilacinum* nonsegmented virus 1) (Herrero, 2016; Zhang et al., 2014; Zhou et al., 2016), (iv) six members of an unclassified taxon of monosegmented, puta-

tive dsRNA viruses (proposed genus “Unirnavirus”) represented by *Beauveria bassiana* RNA virus 1 (also *Alternaria longipes* dsRNA virus 1, *Colletotrichum higginsianum* nonsegmented dsRNA virus 1, *Penicillium janczewskii* B. *bassiana*-like virus 1, *Ustilaginoidea virens* RNA virus M, and *Ustilaginoidea virens* unassigned RNA virus) (Campo et al., 2016; Jiang et al., 2015; Koloniuk et al., 2015; Kotta-Loizou et al., 2015; Lin et al., 2015; Nerva et al., 2015; Zhu et al., 2015), and (v) nine members of an unclassified taxon of bisegmented, putative dsRNA viruses represented by *Curvularia thermal* tolerance virus from *C. protuberata* (also *Cryphonectria parasitica* bipartite mycovirus 1, *Fusarium graminearum* dsRNA mycovirus 4, *Gremmeniella abietina* RNA virus 6, *Heterobasidion* RNA virus 6, *Penicillium aurantiogriseum* bipartite virus 1, *Rhizoctonia fumigata* mycovirus, *Rhizoctonia solani* dsRNA virus 1, and *Sclerotium hydrophilum* virus 1) (Botella et al., 2015; Márquez et al., 2007;



**Fig. 3.** Phylogenetic tree and pairwise identities. (A) Sequences of the ORF2/RdRp translation products were aligned using MUSCLE and then subjected to phylogenetic analysis using PhyML as described in Materials and Methods. Proposed amalgaviruses new to this report are labeled in gray. The tree is displayed as a rectangular phylogram rooted at the midpoint. Branch support values are shown in %, and those with support values <70% are drawn with thinner lines. Scale bar, average number of substitutions per alignment position. See Table S3 for a summary of abbreviations and GenBank numbers. Vertical lines: approved or proposed spans of genera and families (family *Amalgaviridae* has been previously proposed to encompass proposed genus *Zybavirus* by Depierreux et al. (2016) and is newly proposed to encompass proposed genus *Anlovirus* here). (B) Sequences of the ORF1 (lower left) and ORF2 (upper right) translation products of the indicated viruses were compared in pairs using EMBOSS: needle or needleall. Sequence identity scores are shown in %. Gray shading highlights the values within each of the approved or proposed genera.

Nerva et al., 2015; Vainio et al., 2012; Wang et al., 2016; Yu et al., 2009; Zheng et al., 2013). The results from maximum-likelihood (PhyML) phylogenetic analyses, using either MAFFT or MUSCLE for the initial sequence alignment and the LG, rtREV, JTT, or WAG substitution model, provided evidence that a taxon represented by AnloV1/ATCC30860 is distinguishable from those of the other viruses, though much more closely related to those of ZbV-Z and plant amalgaviruses (Fig. 3A; also Fig. S1). Sequence identity scores from pairwise comparisons of the predicted ORF1p sequences and the ORF2 (RdRp) portion of the predicted ORF1+2p sequences of these viruses (Fig. 3B) were consistent with the phylogenetic results and provided further evidence that AnloV1 represents a distinct taxon. We propose that AnloV1/ATCC30860 is the prototype strain of a new species, which we further propose to assign as type species of a new genus in family *Amalgaviridae*. We suggest the name "Anlovirus" for this genus, as derived from the name of the prototype host *Antonospora locustae*.

### 3.5. Sequence of an AnloV1-related virus mined from the TSA database

The predicted AnloV1 ORF1+2p sequence was used for database searches in an effort to find sequences of other related viruses. When performed on the TSA database for arthropods, the

two highest-scoring accessions (GenBank GAXI02024117 and GAXI02024802) were from the giant springtail *Tetrodontophora bielanensis*, a basal hexapod from class/order Collembola. Both of these TSA accessions are relatively long (1606 and 1853 nt) and align with opposite halves of the AnloV1 genome sequence. In fact when examined, the two accessions were found to overlap by 5 nt. When combined into a single contig based on that small overlap, the new length is 3454 nt, approximating that of an amalga-like virus genome. Concerned by the shortness of the preceding overlap, we turned to the SRA data set (SRX314901) from which the TSA accessions from *T. bielanensis* had been assembled, and we there found 1196 reads that matched the 3454-nt contig. Upon assembling these reads, two contigs were obtained, corresponding to GAXI02024117 and GAXI02024802, but slightly longer than before and now overlapping by 29 nt. When combined into a single contig based on that overlap, the new length is 3475 nt due to small extensions at both termini relative to the 3454-nt contig. Only 53 positions in this contig were represented in fewer than 5 of the 1196 reads: 34 positions at one end, 15 positions at the junction of the two assembled contigs, and 4 positions at the other end. At all other positions, coverage depths ranged from 5 to 207. We henceforth refer to this element as *Tetrodontophora bielanensis* associated virus 1 (TebiAV1). The new 3475-nt partial genome sequence for TebiAV1 is provided in supplementary file Data S1.

### 3.6. Primary features and analyses of TebiAV1 genome sequence

The minus strand of TebiAV1 is devoid of long ORFs, including none  $\geq 255$  nt in length between flanking stop codons. The plus strand, on the other hand, includes two long ORFs, which are partially overlapping (Fig. 2). ORF1 spans positions <1–815, extending from the 5' end to a flanking UAA stop codon at positions 816–818 (Table 1). The 5'-most in-frame AUG codon is at positions 114–116, and may or may not represent the functional start codon of ORF1p: the functional start codon may, for example, be further upstream than yet encompassed by the contig. Assuming that the latter is true (as seems likely to us) and that the protein-coding region of ORF1 spans positions <1–815, it is predicted to encode a >271-aa, >32-kDa product (Table 1). ORF2 spans positions 721–3225, extending between a UAG stop codon at positions 718–720 and a UAG stop codon at positions 3226–3228 (Table 1). ORF1 and ORF2 thus overlap by 95 nt, not including the flanking stop codons. Notably, ORF2 is in the +1 frame relative to ORF1, and within the ORF1–ORF2 overlap region is the sequence UUU.CUU.G (underlines, ORF1 codon boundaries) at positions 729–735, identical to the proposed +1 PRF motif in AnloV1. If +1 PRF indeed occurs in this motif (translation of slippage codon UUU/UUC, then UUG), then the resulting ORF1 + ORF2 fusion is predicted to span positions <1–731:733–3325 and to encode a >1074-aa, >125-kDa product that we again designate ORF1+2p (Table 1).

In searches against the full NR database, the predicted ORF1p sequence of TebiAV1 yielded no significant hits (all E-values  $\geq 1.1$ ). The predicted ORF1+2p sequence, on the other hand, yielded significant hits with viral RdRp sequences, those of ZbV-Z and plant amalgaviruses constituting the top 11 hits (E-values, 1e–36 to 6e–11). The ORF2 (RdRp) portion of the predicted ORF1+2p sequence of TebiAV1 was additionally included in the same phylogenetic analyses as described above for AnloV1, the results of which provided evidence that TebiAV1 is most closely related to AnloV1, in a shared clade that diverges between ZbV-Z and plant amalgaviruses in the phylogram (Fig. 3A; also Fig. S1). Sequence identity scores from pairwise comparisons of the predicted ORF1p sequences and the ORF2 (RdRp) portion of the predicted ORF1+2p sequences of these viruses (Fig. 3B) were consistent with the phylogenetic results and provided further evidence that TebiAV1 represents a distinct taxon that it appears to share with AnloV1. We conclude that TebiAV1 is the prototype strain of another new species, which we propose to place in proposed genus “Anlovirus” along with AnloV1.

## 4. Discussion

In ZbV-Z and most plant amalgaviruses (Depierreux et al., 2016; Nibert et al., 2016), as originally shown in influenza A viruses (Firth et al., 2012; Jagger et al., 2012), the consensus motif for +1 PRF can be identified as UUU.CGN.N (underlines: codon boundaries of the upstream ORF). Within this motif, +1 slippage of the AAG anticodon is thought to occur in the ribosomal P-site such that after translation of the slippage codon UUU/UUC, the next translated codon is GNN. The presence of a rare Arg codon, CGN, in this motif is thought to stimulate +1 slippage consequent to slow decoding of this codon in the ribosomal A-site (Firth et al., 2012). In the apparent motif for +1 PRF in AnloV1 and TebiAV1, the rare Arg codon is replaced by the Leu codon CUU, suggesting that the core motif for +1 PRF is more simply UUU.C, as indeed previously anticipated by Firth et al. (2012). Other cis-acting sequences or factors involved in regulating +1 PRF, beyond the UUU.C motif in AnloV1 and TebiAV1, remain to be determined.

Notably, 12% of the EST clones with usable sequences obtained from *A. locustae* ATCC30860 spores by Williams et al. (2005) map

to the AnloV1 genome, accounting for almost half of the EST accessions from that study that originally lacked discernible homologs. AnloV1-derived clones were thus highly represented in the EST library. Our capacity to visualize AnloV1 dsRNA as a strong band on agarose gels (see Fig. 1) provides further evidence that *A. locustae* ATCC30860 spores contain large amounts of this viral RNA. Moreover, two different purified preparations of these spores, obtained from the same commercial source more than 10 years apart, are represented by the current results. We speculate that AnloV1 may make use of one or more mechanism for determining its efficient partitioning into *A. locustae* spores, thereby promoting its vertical transmission.

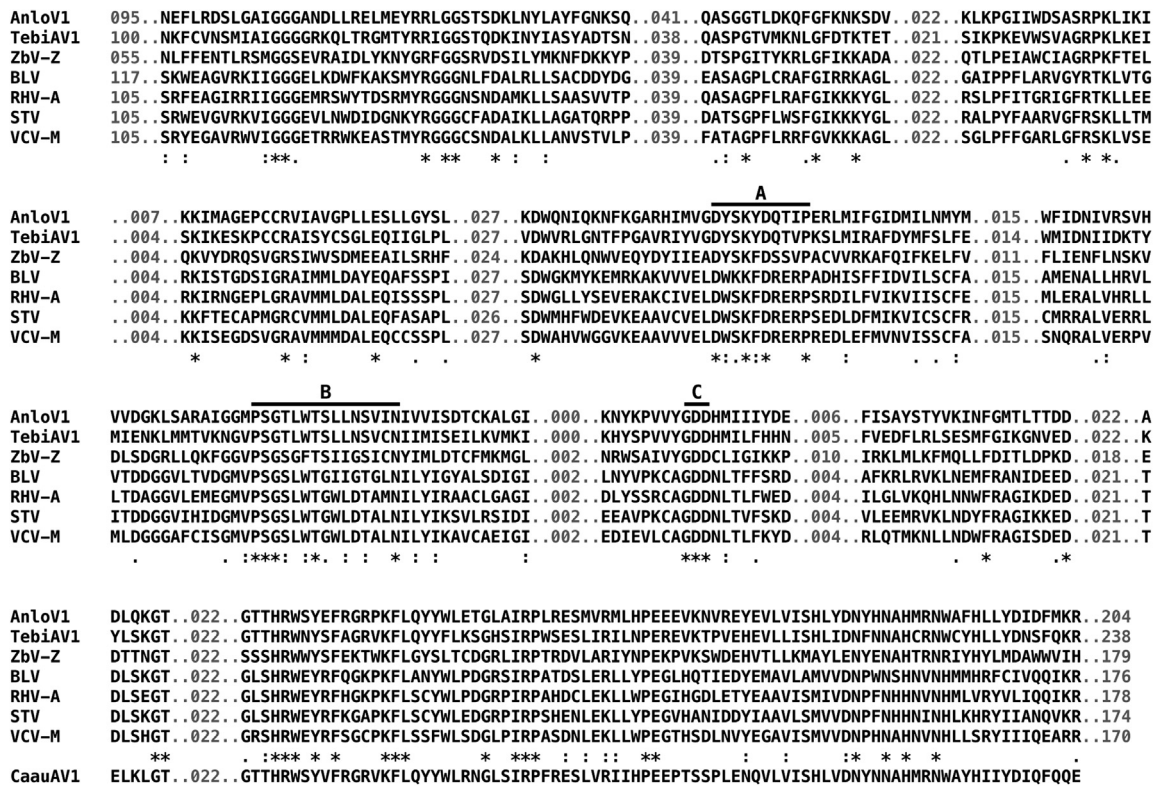
The transcriptome of *T. bielanensis* was deposited in GenBank (BioProject PRJNA219607) as part of the larger 1KITE (1000 Insect Transcriptome Evolution) project (Misof et al., 2014). The analyzed specimens of *T. bielanensis* (30 individuals) were collected from the wild near Ostritz, Germany. Because RNA extraction for transcriptome analysis was then performed on the whole organisms (Misof et al., 2014), that extract would have also contained RNA from any *T. bielanensis* symbionts (e.g., parasites) that may have been present in those specimens. That is why we identify TebiAV1 as a *T. bielanensis* “associated” virus, not a *T. bielanensis* virus per se. The relatively small fraction of SRA reads that map to the TebiAV1 contig, 0.009% (Table S1), may be consistent with this explanation. Moreover, given documented associations between springtails and microsporidia (Bigliardi and Carapelli, 2002; Purrini and Weiser, 1983; Weiser and Purrini, 1980), as well as the similarity of TebiAV1 to AnloV1, we speculate that TebiAV1 may have derived from an unidentified microsporidian, or related organism, that was symbiotically associated with one or more of the analyzed specimens of *T. bielanensis*. In fact, by choosing the 23 GenBank-deposited *A. locustae* protein sequences with lengths >600 aa for use as representative queries in tblastn searches of the TSA database for *T. bielanensis*, we were able to identify 4 strong (E=0.0) hits that, when translated and used to search the full NR database, scored topmost hits (again E=0.0) to protein homologs from different basal fungi (Table 2). We therefore consider it likely that at least one such organism was associated with the analyzed specimens of *T. bielanensis* and served as the direct host for TebiAV1.

When the AnloV1 ORF1+2p sequence was used for the original tblastn search of the TSA database for arthropods, as described in Results, the next highest-scoring accession below the two from *T. bielanensis* was from the two-pronged bristletail *Campodea augens* (GenBank GAYN02026491), another basal hexapod, though in this case from class/order Diplura. As in the case of *T. bielanensis*, the transcriptome of *C. augens* was deposited in GenBank (BioProject PRJNA219535) as part of the larger 1KITE project. The identified sequence from *C. augens* is small, only 266 nt, but by searching the associated SRA data set (SRX314832), we were able to extend this contig to 440 nt (see supplementary file Data S1), which is notable for containing an end-to-end ORF. When the 146-aa translation product was then used for pairwise comparisons, it was found to be 59% and 62% identical to the ORF2/RdRp products of AnloV1 and TebiAV1, respectively. Moreover, this partial translation product clustered with those of AnloV1 and TebiAV1 in phylogenetic analyses (Fig. 3A; also Fig. S1). These findings thus represent tentative evidence for another member, *Campodea augens* associated virus 1 (CaauAV1), of proposed genus “Anlovirus”.

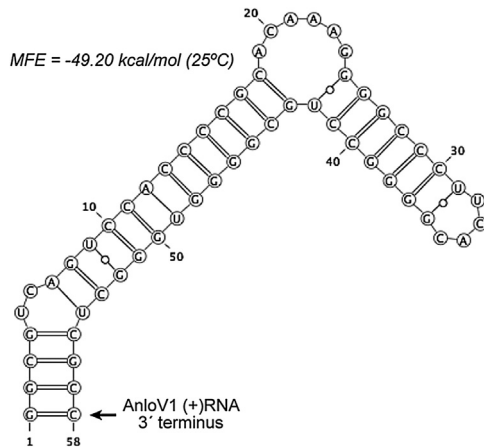
The specific nature and function of protein ORF1p of plant amalgaviruses and ZbV-Z remain unknown. Although still possible that ORF1p of these viruses is an icosahedral capsid-forming protein (Depierreux et al., 2016; Schmitt and Neuhausen, 1994), it seems more likely to have some other form and function, in part because (i) isometric virus-like particles have failed to be seen in amalgavirus-infected cells or purified from them (Isogai et al., 2011; Liu and Chen, 2009; Martin et al., 2011; Sabanadzovic et al., 2009, 2010) and







**Fig. 5.** ORF2p (RdRp) alignment. Multiple sequence alignments were generated for the sequences derived from ORF2 that are predicted to be present in ORF1+2p of each virus. Blocks of sequences that were identically aligned by PROMALS3D, MAFFT, and MUSCLE are shown in black; whereas sequences less consistently aligned by these programs are shown in gray as the number of residues between aligned blocks. Consensus symbols were assigned according to the BLOSUM62 scoring matrix: \*, identical;., nonidentical with all pairwise scores  $\geq 1$ ; .. nonidentical with all pairwise scores  $\geq 0$  and at least one pairwise score = 0. The partial sequence from CaaAV1 has been manually aligned at the bottom. Conserved RdRp motifs A, B, and C are overlined and labeled.



**Fig. 6.** Predicted secondary structure at 3' end of AnLoV1 plus-strand (+) RNA. RNAfold was used to predict RNA secondary structures formed by progressively larger 3'-terminal portions of the AnLoV1 (+)RNA sequence. The illustrated structure is the optimal one predicted to form when the 3'-terminal 58–180 nt are tested. If the complete AnLoV1 (+)RNA sequence is tested, or 3'-terminal portions of it  $> 180$  nt in length, the G at position 1 in this figure is predicted to be recruited into another, upstream structure, so that the 3'-terminal C residue of the AnLoV1 (+)RNA (nt 58 in this figure) is predicted to remain unpaired.

of over 220 invertebrate species. Among these invertebrate-associated viruses are two that fall within the span of family *Amalgaviridae* proposed by Depierreux et al. (2016): Hubei partiti-like virus 59 (HPLV59) (GenBank KX884149) and Beihai barnacle virus 14 (BBV14) (GenBank KX884071), from the respective transcriptomes of an insect (dipteran) and a crustacean (barnacle) species. Based on new phylogenetic analyses, it appears that

HPLV59 might be best assigned to proposed genus Anlovirus and BBV14 to proposed genus Zybavirus (Fig. S2A). HPLV59 and BBV14 are also notable for containing the +1 PRF motif UUU.CUU.G properly positioned within their ORF1–ORF2 overlap regions; for having a central region with strong coiled coil propensity in ORF1p; and for sharing the two newly suggested amalgavirus signature motifs (see above) in the RdRp region of ORF1+2p (Fig. S2B). Thus, all new amalga-like viruses discussed in this report share a number of common features with plant amalgaviruses and Zbv-Z.

**Acknowledgements**

We thank Lee Anne Merrill (President of M&R Durango, Inc., Bayfield, CO, USA) for providing purified *A. locustae* spores for use in this study. J.D.P. completed his work on this project during a lab rotation for the Ph.D. Training Program in Virology at Harvard University, Cambridge, MA, USA, and was supported in part by NIH grant 2T32AI007245-31. P.J.K. is a Senior Fellow of the Canadian Institute for Advanced Research and was supported in part by Canadian Institutes of Health Research grant MOP-42517. M.L.N. was supported in part by a subcontract from NIH grant 5R01GM033050-33.

**Appendix A. Supplementary data**

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.virusres.2017.02.015>.

## References

- Bigliardi, E., Carapelli, A., 2002. Microsporidia in the springtail *Isotomurus fucicolus* (Collembola, Isotomidae) and possible pathways of parasite transmission. *Ital. J. Zool.* 69, 109–113.
- Botella, L., Vainio, E.J., Hantula, J., Diez, J.J., Jankovsky, L., 2015. Description and prevalence of a putative novel mycovirus within the conifer pathogen *Gremmeniella abietina*. *Arch. Virol.* 160, 1967–1975.
- Bruenn, J.A., 2003. A structural and primary sequence comparison of the viral RNA-dependent RNA polymerases. *Nucleic Acids Res.* 31, 1821–1829.
- Campo, S., Gilbert, K.B., Carrington, J.C., 2016. Small RNA-based antiviral defense in the phytopathogenic fungus *Colletotrichum higginsianum*. *PLoS Pathog.* 12, e1005640.
- Castillo, A., Cottet, L., Castro, M., Sepúlveda, F., 2011. Rapid isolation of mycoviral double-stranded RNA from *Botrytis cinerea* and *Saccharomyces cerevisiae*. *Virol. J.* 8, 38.
- Coutts, R.H.A., Livieratos, I.C., 2003. A rapid method for sequencing the 50- and 30-termini of dsRNA viral templates using RLM-RACE. *J. Phytopathol.* 151, 525–527.
- DePaulo, J.J., Powell, C.A., 1995. Extraction of double-stranded RNA from plant tissues without the use of organic solvents. *Plant Dis.* 79, 246–248.
- Delorenzi, M., Speed, T., 2002. An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics* 18, 617–625.
- Depierreux, D., Vong, M., Nibert, M.L., 2016. Nucleotide sequence of *Zygosaccharomyces bailii* virus Z: Evidence for +1 programmed ribosomal frameshifting and for assignment to family *Amalgaviridae*. *Virus Res.* 217, 115–124.
- Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.
- Firth, A.E., Jagger, B.W., Wise, H.M., Nelson, C.C., Parsawar, K., Wills, N.M., Napthine, S., Taubenberger, J.K., Digard, P., Atkins, J.F., 2012. Ribosomal frameshifting used in influenza A virus expression occurs within the sequence UCCUUU.CGU and is in the +1 direction. *Open Biol.* 2, 120109.
- Gruber, A.R., Lorenz, R., Bernhart, S.H., Neuböck, R., Hofacker, I.L., 2008. The Vienna RNA website. *Nucleic Acids Res.* 36 (Web Server issue), W70–74.
- Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321.
- Herrero, N., 2016. A novel monopartite dsRNA virus isolated from the entomopathogenic and nematophagous fungus *Purpureocillium lilacinum*. *Arch. Virol.* 161, 3375–3384.
- Huang, X., Madan, A., 1999. CAP3: a DNA sequence assembly program. *Genome Res.* 9, 868–877.
- Isogai, M., Nakamura, T., Ishii, K., Watanabe, M., Yamagishi, N., Yoshikawa, N., 2011. Histochemical detection of Blueberry latent virus in highbush blueberry plant. *J. Gen. Plant Pathol.* 77, 304–306.
- Issi, I.V., 2002. Parasitic systems of Microsporidia: descriptions and terminology questions. *Parazitologia* 36, 478–492.
- Jagger, B.W., Wise, H.M., Kash, J.C., Walters, K.A., Wills, N.M., Xiao, Y.L., Dunfee, R.L., Schwartzman, L.M., Ozinsky, A., Bell, G.L., Dalton, R.M., Lo, A., Efstathiou, S., Atkins, J.F., Firth, A.E., Taubenberger, J.K., Digard, P., 2012. An overlapping protein-coding region in influenza A virus segment 3 modulates the host response. *Science* 337, 199–204.
- Jiang, Y., Zhang, T., Luo, C., Jiang, D., Li, G., Li, Q., Hsiang, T., Huang, J., 2015. Prevalence and diversity of mycoviruses infecting the plant pathogen *Ustilagoidea vires*. *Virus Res.* 195, 47–56.
- Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780.
- Keeling, P.J., Slamovits, C.H., 2004. Simplicity and complexity of microsporidian genomes. *Eukaryot. Cell* 3, 1363–1369.
- Koloniuk, I., Hrabáková, L., Petrzik, K., 2015. Molecular characterization of a novel amalgavirus from the entomopathogenic fungus *Beauveria bassiana*. *Arch. Virol.* 160, 1585–1588.
- Kotta-Loizou, I., Sipkova, J., Coutts, R.H.A., 2015. Identification and sequence determination of a novel double-stranded RNA mycovirus from the entomopathogenic fungus *Beauveria bassiana*. *Arch. Virol.* 160, 873–875.
- Krupovic, M., Dolja, V.V., Koonin, E.V., 2015. Plant viruses of the *Amalgaviridae* family evolved via recombination between viruses with double-stranded and negative-strand RNA genomes. *Biol. Direct* 10, 12.
- Lange, C.E., 2005. The host and geographical range of the grasshopper pathogen *Paranosema (Nosema) locustae* revisited. *J. Orthoptera Res.* 14, 137–141.
- Lin, Y., Zhang, H., Zhao, C., Liu, S., Guo, L., 2015. The complete genome sequence of a novel mycovirus from *Alternaria longipes* strain HN28. *Arch. Virol.* 160, 577–580.
- Liu, W., Chen, J., 2009. A double-stranded RNA as the genome of a potential virus infecting *Vicia faba*. *Virus Genes* 39, 126–131.
- Liu, H., Fu, Y., Xie, J., Cheng, J., Ghabrial, S.A., Li, G., Yi, X., Jiang, D., 2012. Discovery of novel dsRNA viral sequences by in silico cloning and implications for viral diversity, host range and evolution. *PLoS One* 7, e42147.
- Márquez, L.M., Redman, R.S., Rodriguez, R.J., Roossinck, M.J., 2007. A virus in a fungus in a plant: three-way symbiosis required for thermal tolerance. *Science* 315, 513–515.
- Martin, R.R., Zhou, J., Tzanetakis, I.E., 2011. Blueberry latent virus: an amalgam of the *Partitiviridae* and *Totiviridae*. *Virus Res.* 155, 175–180.
- McDonnell, A.V., Jiang, T., Keating, A.E., Berger, B., 2006. *Paircoil2: improved prediction of coiled coils from sequence*. *Bioinformatics* 22, 356–358.
- Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B., Ware, J., Flouri, T., Beutel, R.G., Niehuis, O., Petersen, M., Izquierdo-Carrasco, F., Wappler, T., Rust, J., Aberer, A.J., Aspöck, U., Aspöck, H., Bartel, D., Blanke, A., Berger, S., Böhm, A., Buckley, T.R., Calcott, B., Chen, J., Friedrich, F., Fukui, M., Fujita, M., Greve, C., Grobe, P., Gu, S., Huang, Y., Jeremiin, L.S., Kawahara, A.Y., Krogmann, L., Kubiak, M., Lanfear, R., Letsch, H., Li, Y., Li, Z., Li, J., Lu, H., Machida, R., Mashimo, Y., Kapli, P., McKenna, D.D., Meng, G., Nakagaki, Y., Navarrete-Heredia, J.L., Ott, M., Ou, Y., Pass, G., Podsiadlowski, L., Pohl, H., von Reumont, B.M., Schütte, K., Sekiya, K., Shimizu, S., Slipinski, A., Stamatakis, A., Song, W., Su, X., Szucsich, N.U., Tan, M., Tan, X., Tang, M., Tang, J., Timelthaler, G., Tomizuka, S., Trautwein, M., Tong, X., Uchifune, T., Walz, M.G., Wiegmann, B.M., Wilbrandt, J., Wipfler, B., Wong, T.K., Wu, Q., Wu, G., Xie, Y., Yang, S., Yang, Q., Yeates, D.K., Yoshizawa, K., Zhang, Q., Zhang, R., Zhang, W., Zhang, Y., Zhao, J., Zhou, C., Zhou, L., Ziesmann, T., Zou, S., Li, Y., Xu, X., Zhang, Y., Yang, H., Wang, J., Wang, J., Kjer, K.M., Zhou, X., et al., 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science* 346, 763–767.
- Morozov, A.Y., Robin, C., Franc, A., 2007. A simple model for the dynamics of a host-parasite-hyperparasite interaction. *J. Theor. Biol.* 249, 246–253.
- Nerva, L., Ciuffo, M., Vallino, M., Margaria, P., Varese, G.C., Gnani, G., Turina, M., 2015. Multiple approaches for the detection and characterization of viral and plasmid symbionts from a collection of marine fungi. *Virus Res.* 219, 22–38.
- Nibert, M.L., Pyle, J.D., Firth, A.E., 2016. A +1 ribosomal frameshifting motif prevalent among plant amalgaviruses. *Virology* 498, 201–208.
- Pei, J., Kim, B.H., Grishin, N.V., 2008. PROMALS3D: a tool for multiple sequence and structure alignment. *Nucleic Acids Res.* 36, 2295–2300.
- Ponty, Y., Leclerc, F., 2015. Drawing and editing the secondary structure(s) of RNA. *Methods Mol. Biol.* 1269, 63–100.
- Purrini, K., Weiser, J., 1983. *Octospora collembolae* n. sp. (Microsporidia, Microspora): a new microsporidian parasite of springtail *Onychiurus quadricellatus* (Onychiuridae:Collembolae). *J. Invertebr. Pathol.* 42, 135–142.
- Sabanadzovic, S., Valverde, R.A., Brown, J.K., Martin, R.R., Tzanetakis, I.E., 2009. Southern tomato virus: the link between the families *Totiviridae* and *Partitiviridae*. *Virus Res.* 140, 130–137.
- Sabanadzovic, S., Abou Ghanem-Sabanadzovic, N., Valverde, R.A., 2010. A novel monopartite dsRNA virus from rhododendron. *Arch. Virol.* 155, 1859–1863.
- Schäffer, A.A., Aravind, L., Madden, T.L., Shavirin, S., Spouge, J.L., Wolf, Y.I., Koonin, E.V., Altschul, S.F., 2001. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res.* 29, 2994–3005.
- Schmitt, M.J., Neuhausen, F., 1994. Killer toxin-secreting double-stranded RNA mycoviruses in the yeasts *Hanseniaspora uvarum* and *Zygosaccharomyces bailii*. *J. Virol.* 68, 1765–1772.
- 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., Buchmann, J., Wang, W., Xu, J., Holmes, E.C., Zhang, Y.Z., 2016. Redefining the invertebrate RNA virosphere. *Nature*. <http://dx.doi.org/10.1038/nature20167> (in press).
- Slamovits, C.H., Williams, B.A., Keeling, P.J., 2004. Transfer of *Nosema locustae* (Microsporidia) to *Antonospora locustae* n. comb. based on molecular and ultrastructural data. *J. Eukaryot. Microbiol.* 51, 207–213.
- Stentiford, G.D., Becnel, J.J., Weiss, L.M., Keeling, P.J., Didier, E.S., Williams, B.A., Bjornson, S., Kent, M.L., Freeman, M.A., Brown, M.J., Troemel, E.R., Roesel, K., Sokolova, Y., Snowden, K.F., Solter, L., 2016. Microsporidia—emergent pathogens in the global food chain. *Trends Parasitol.* 32, 336–348.
- Vainio, E.J., Hyder, R., Aday, G., Hansen, E., Piri, T., Doğmuş-Lehtijärvi, T., Lehtijärvi, A., Korhonen, K., Hantula, J., 2012. Population structure of a novel putative mycovirus infecting the conifer root-rot fungus *Heterobasidion annosum* sensu lato. *Virology* 422, 366–376.
- Wang, C., Wu, J., Zhu, X., Chen, J., 2016. Complete nucleotide sequences of dsRNA2 and dsRNA7 detected in the phytopathogenic fungus *Sclerotium hydrophilum* and their close phylogenetic relationship to a group of unclassified viruses. *Virus Genes* 52, 823–827.
- Weiser, J., Purrini, K., 1980. Seven new microsporidian parasites of springtails (Collembola) in the Federal Republic of Germany. *Z. Parasitenk.* 62, 75–84.
- Williams, B.A., Slamovits, C.H., Patron, N.J., Fast, N.M., Keeling, P.J., 2005. A high frequency of overlapping gene expression in compacted eukaryotic genomes. *Proc. Natl. Acad. Sci. U. S. A.* 102, 10936–10941.
- Yu, J., Kwon, S.J., Lee, K.M., Son, M., Kim, K.H., 2009. Complete nucleotide sequence of double-stranded RNA viruses from *Fusarium graminearum* strain DK3. *Arch. Virol.* 154, 1855–1858.
- Zhang, T., Jiang, Y., Dong, W., 2014. A novel monopartite dsRNA virus isolated from the phytopathogenic fungus *Ustilagoidea vires* and ancestrally related to a mitochondria-associated dsRNA in the green alga *Bryopsis*. *Virology* 462, 227–235.
- Zheng, L., Liu, H., Zhang, M., Cao, X., Zhou, E., 2013. The complete genomic sequence of a novel mycovirus from *Rhizoctonia solani* AG-1 IA strain B275. *Arch. Virol.* 158, 1609–1612.
- Zhou, Q., Zhong, J., Hu, Y., Da Gao, B., 2016. A novel nonsegmented double-stranded RNA mycovirus identified in the phytopathogenic fungus *Nigrospora oryzae* shows similarity to partitivirus-like viruses. *Arch. Virol.* 161, 229–232.
- Zhu, H.J., Chen, D., Zhong, J., Zhang, S.Y., Gao, B.D., 2015. A novel mycovirus identified from the rice false smut fungus *Ustilagoidea vires*. *Virus Genes* 51, 159–162.