Genomics and proteomics of Apis mellifera filamentous virus isolated from honeybees in China

Dahe Yang\textsuperscript{a,b,1}, Jun Wang\textsuperscript{b,1}, Xi Wang\textsuperscript{b}, Fei Deng\textsuperscript{b}, Qingyun Diao\textsuperscript{a}, Manli Wang\textsuperscript{b}, Zhihong Hu\textsuperscript{b,*, c}, Chunsheng Hou\textsuperscript{c,∗}

\textsuperscript{a} Institute of Apicultural Research, Chinese Academy of Agricultural Sciences, Beijing, 100093, China
\textsuperscript{b} State Key Laboratory of Virology and National Virus Resource Center, Wuhan Institute of Virology, Center for Biosafety Mega-Science, Chinese Academy of Sciences, Wuhan, 430071, China
\textsuperscript{c} Institute of Bast Fiber Crops, Chinese Academy of Agricultural Sciences, Changsha, 410205, China

\textbf{A B S T R A C T}

Apis mellifera filamentous virus (AmFV) is a large DNA virus that is endemic in honeybee colonies. The genome sequence of the AmFV Swiss isolate (AmFV CH–C05) has been reported, but so far very few molecular studies have been conducted on this virus. In this study, we isolated and purified AmFV (AmFV CN) from Chinese honeybee (Apis mellifera) colonies and elucidated its genomics and proteomics. Electron microscopy showed ovoid purified virions with dimensions of 300–500 × 210–285 nm, wrapping a 3165 × 40 nm filamentous nucleocapsid in three figure-eight loops. Unlike AmFV CH–C05, which was reported to have a circular genome, our data suggest that AmFV CN has a linear genome of approximately 493 kb. A total of 197 ORFs were identified, among which 36 putative genes including 18 baculoviral homologs were annotated. The overall nucleotide similarity between the CN and CH–C05 isolates was 96.9%. Several ORFs were newly annotated in AmFV CN, including homologs of \textit{per os} infectivity factor 4 (PIF4) and a putative integrase. Phylogenomic analysis placed AmFVs on a separate branch within the newly proposed virus class Naldaviricetes. Proteomic analysis revealed 47 AmFV virion-associated proteins, of which 14 had over 50% sequence coverage, suggesting that they are likely to be main structural proteins. In addition, all six of the annotated PIFs (PIF-0–5) were identified by proteomics, suggesting that they may function as entry factors in AmFV infection. This study provides fundamental information regarding the molecular biology of AmFV.

1. Introduction

Honeybees are indispensable pollinators in natural ecosystems and contribute to the production of approximately 70% of the crops used for human consumption (Klein \textit{et al.}, 2007). However, multiple factors threaten honeybees, such as pesticides, and parasites including viruses (Brosi \textit{et al.}, 2017). To date, more than 30 honeybee viruses have been reported, most of which are RNA viruses (McMenamin and Genersch, 2015; Remnant \textit{et al.}, 2017; Beaurepaire \textit{et al.}, 2020). Apis mellifera filamentous virus (AmFV) is one of the few DNA viruses identified in honeybees and has been studied less than the RNA viruses.

AmFV was initially reported as a honeybee pathogen in the United States in 1978 (Clark, 1978). Acutely infected bees become weak and gather at the hive entrance, while severely infected honeybees exhibit milky-white hemolymph due to tissue degradation (Clark, 1978). In general, AmFV appears to be a weak pathogen, but is endemic in honeybee colonies. For example, it is considered to be the most common and least harmful bee virus in Britain (Bailey, 1982). While AmFV appears to be a weak pathogen, it is not unreasonable to suggest that it may weaken the bee to an extent that makes it more susceptible to other pathogens. Initially, the presence of AmFV was diagnosed using electron microscopy. AmFV was first reported to be an ellipsoidal (400 × 100 nm), enveloped virus with a long filamentous nucleocapsid (3060 × 60 nm) (Clark, 1978). Later, it was characterized as a DNA virus of slightly different size (450 × 150 nm and 3000 × 40 nm for virion and nucleocapsid, respectively) (Bailey \textit{et al.}, 1981). The nucleocapsid morphology of AmFV is unique in that it forms three figure-eight loops inside the envelope (Sitaropoulouab \textit{et al.}, 1989).
Partial sequences of AmFV were first derived from genome sequencing of *Varroa destructor* mites, but at the time they were referred to as baculovirus-related (Cornman et al., 2010). A breakthrough in AmFV research was made in 2015, when the complete viral genome was sequenced from infected worker honeybees collected in Switzerland (Gauthier et al., 2015); this led to the molecular detection of AmFV in practice.

The prevalence of AmFV has been surveyed in the USA, Switzerland, France, Sweden, China, Syria, Czech Republic, and Argentina, showing that it is commonly found worldwide (Gauthier et al., 2015; Hartmann et al., 2015; Hou et al., 2016; Abou Kubaa, 2018; Prodelalova et al., 2019; Quintana, 2019). AmFV is detectable year-round, with higher viral copy numbers found in the Spring (Hartmann et al., 2015). Apart from honeybees, AmFVs have been detected in a wide range of solitary bee species (Ravoet et al., 2014) and even in honey (Bovo et al., 2018, 2020).

The double-stranded DNA genome of the Swiss strain AmFV (AmFV CH–C05) is approximately 496 kb, encoding 247 open reading frames (ORFs). However, only a small portion (~16%) of the predicted ORFs have been annotated, including 13 homologs of baculovirus genes, including per as infectivity factors (PIFs) and baculovirus repeated ORFs (BROs) (Gauthier et al., 2015). Apart from genome sequencing, very few molecular studies have been conducted on AmFV. An early biochemistry study revealed that there were 12 AmFV structural proteins (Bailey et al., 1981), but to our knowledge, no further relevant investigation has been performed since then. Molecular characterization of AmFV is indispensable to better understand its pathogenicity and its interaction with its host. In this study, we isolated and purified AmFV virions from honeybee colonies in China and conducted next-generation genomic sequencing and compared it to that of the Swiss strain. We performed proteomics to identify virion-associated proteins.

### 2. Materials and methods

#### 2.1. Virus purification and viral DNA extraction

*Apis mellifera* naturally infected with AmFV were collected from Henan Province, China. The AmFV infection was detected by PCR. Briefly, total DNA was extracted from bee workers by phenol-chloroform and ethanol precipitation. PCR amplification was performed using AmFV specific primers (5′-CAGAGATTCCGTTTTGTGAGT-3′ and 5′-CATGGTGCCGAAATCTGGCT-3′) (Gauthier et al., 2015). The identity of the PCR products was further confirmed by Sanger sequencing. Virions (AmFV CN) were purified from approximately 40 honeybee adults as previously described (Bailey et al., 1981; Laughton and Siva-Jothy, 2011) with slight modifications. Briefly, bees were homogenized in extraction buffer (0.01 mol/L ammonium acetate, 0.02% diethylidithiocarbamate, and 0.01% Triton X-100). The homogenate was filtered using 4-layer gauze, then centrifuged at 1000 g for 30 min. The supernatant was centrifuged on a sucrose density gradient (20%–60%) at 40000 g for 1 h. The band at 50% sucrose was collected and purified to remove sucrose. The purified virions were imaged by transmission electron microscopy (TEM) using a 100 kV Hitachi H-7000FA microscope. DNA extraction was performed as previously described (Gauthier et al., 2015).

#### 2.2. Genome sequencing and bioinformatic analysis

Virion DNA was sequenced using the Illumina Hiseq 3000 System with shotgun strategy at the Sequencing Platform of the National Key Laboratory of Crop Genetic Improvement at Huazhong Agricultural University (Wuhan, China). The reads were quality controlled and pre-processed using Trimmomatic (version 0.32), then assembled with Trinity (version 2.5.1). Gaps and unreadable sequences were amplified by PCR and confirmed by Sanger sequencing. ORFs were identified using the FGENESV program (http://linux1.softberry.com/berry.phtml) and ORFFinder (http://www.ncbi.nlm.nih.gov/orffinder/), adopting the criteria of polypeptide length >100, standard ATG start codon, and minimal overlap. A genome map was constructed using an in-house Python script. Gene annotation and function prediction were performed using the NCBI BLASTP algorithm (https://blast.ncbi.nlm.nih.gov/Blast.cgi) against the nr protein and UniRef90 (https://www.uniprot.org/large) databases. Conserved domains were determined using RPS-BLAST with the Conserved Domain Database (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) and the hmmscan to Pfam database, with a minimum e-value of 1.0e-3 in both cases. The annotated genome sequence data have been submitted to GenBank (http://www.ncbi.nlm.nih.gov/genbank) under accession number OK392616. The genome sequence of AmFV CH–C05 was used as a reference for comparison (GenBank accession number NC_027925.1, Gauthier et al., 2015).

#### 2.3. Phylogenetic analysis

All PIF sequences were aligned using ClustalW with MEGA6 by using default settings. Phylogenetic analysis was conducted using the concatenated PIF amino-acid sequences. The phylogenetic tree was constructed using MEGA6 with the substitution model (LG + G + I) by using the maximum-likelihood method with 1000 bootstrap replicates.

#### 2.4. Proteomics

Shotgun proteomics was used to identify AmFV virion-associated proteins. Briefly, purified AmFV virions were suspended, reduced, alkylated, and subjected to in-solution trypsin digestion. Digested peptides were subjected to liquid chromatography-tandem mass spectrometry (LC-MS/MS) using Q-Exactive Plus coupled to an Easy nLC 1200 (Thermo Fisher Scientific). Proteomics were performed by Bioprofile (Shanghai, China). Peptide sequences were analyzed using the UniProt Protein Database (https://www.uniprot.org/) and Conserved Domain Database (https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml; Lu et al., 2000) with an E-value < 0.1.

### 3. Results

#### 3.1. AmFV purification and morphology

AmFVs were purified from naturally infected honeybees collected in Henan Province, China. Transmission electron micrographs showed the virions were ovoid in shape with a size of approximately 300–500 × 210–285 nm (Fig. 1A), wrapping a long filamentous nucleoprotein into three superimposed figure-eight loops (Fig. 1B). The size of the filamentous nucleoprotein was approximately 3165 × 40 nm. The morphology we observed was similar to those reported in previous studies (Clark, 1978; Bailey et al., 1981; Sitorpoulou et al., 1989).

#### 3.2. Genome overview

The AmFV CN genome was assembled from 37,330,706 high-quality reads with an average coverage of 7576 ×. There were a few gaps (total of ~2843 bp) that could not be filled by PCR and Sanger sequencing, likely due to the complexity of the DNA sequence and/or structure in those regions. Likewise, gaps of approximately 2500 bp also exist in the reported AmFV CH–C05 genome (Gauthier et al., 2015). The size of the assembled genome was 492,752 bp, which was 3644 bp shorter than that of CH–C05 (~496,396 bp). The overall identity between AmFV CN and CH–C05 was 96.9%. The G + C content of AmFV CN was 50.6%, similar to the 50.9% of CH–C05 (Gauthier et al., 2015). However, unlike the CH–C05 genome, which was reported to be circular (Gauthier et al., 2015), our data indicate that the AmFV CN has a linear genome, with no evidence of overlap between the two ends during genome assembly.

Initially, 157 methionine-initiated ORFs with a minimum length of 100 residues were predicted in AmFV CN. When the genome of AmFV CH–C05 (Gauthier et al., 2015) was used as the reference, an additional 40
ORFs with lengths shorter than 100 residues were identified in AmFV CN, increasing the total ORF number to 197 (Fig. 2). Coding regions constituted 63% of the AmFV CN genome, similar to that observed in the CH-05 genome (65%). In comparison to CH-C05 (a total of 241 ORFs), 45 ORFs were missing in AmFV CN, of which 32 were shorter than 100 residues (Table 1). For consistency, the ORFs in AmFV CN were named from their homologs in CH-05 (NC_027925.1). Notably, ORF183 of CH-C05 was split into two separate ORFs in AmFV CN, which were designated AmFV_183 and AmFV_183a. The genome organization between CN and CH-C05 is highly conserved.

The ORFs were annotated on the basis of homology. Thirty-six ORFs had homologs in the genomes of other species (from viruses, eukaryotes, and bacteria) in public sequence databases (Table 2). Based on predicted functions from databases, these ORFs were categorized as: 6 potential DNA replication and nucleotide metabolism, 6 PIFs, 9 BROs, and 15 with putative other or unknown functions. All 36 annotated ORFs in AmFV CN...
were present in AmFV CH-C05 with high protein sequence identities (>90%, Table 2).

As a large DNA virus, AmFV CN has several key factors that facilitate DNA replication. AmFV_074 encodes a hypothetical type-B DNA polymerase with a predicted length of 1960 residues. AmFV_095 is a putative integrase and DNA ligase, respectively, which are likely involved in viral DNA replication. Like many DNA viruses, AmFV also encodes enzymes involved in nucleotide metabolism. AmFV_023, encoding a putative AAAþ ATPase, was also identified in the AmFV CN genome sequence (Gauthier et al., 2015), but are noted in the updated genome sequence in GenBank (NC_027925.1). Bros were first identified in the genome of Lymantria dispar nucleopolyhedrovirus (LdMNPV), which contains 16 repeated ORFs (Kuzio et al., 1999). The number of BROs in AmFV CN is in the same range of the numbers present in baculoviruses, 0 to 16 (Li et al., 2021). BROs constitute a superfamily identified in invertebrate dsDNA viruses, bacteriophages, and bacteria that contain a conserved N-terminal predicted DNA-binding motif (Jakob et al., 2001; Bideshi et al., 2003). Their function(s) is not clear, although several studies have shown that some have DNA-binding activities and interact with laminin or translation-associated-proteins (Zemskov et al., 2000; Kang et al., 2003; Kotani et al., 2015). The multiple BROs found in AmFV support the hypothesis that BROs play important roles in viral interactions with invertebrates (Bideshi et al., 2003).

In addition to the above genes, 15 putative genes were identified with homology to genes from other species (Table 2).

### 3.3. Phylogeny analysis based on PIFs

Of the 36 annotated genes of AmFV CN, 18 (including six PIFs, nine BROs, DNA polymerase, RR1, and RR2) have homologs in baculoviruses, suggesting that it is a baculo-like virus. Based on shared PIFs, the new virus class *Naldaviricetes* was recently proposed by the International Committee on Taxonomy of Viruses (ICTV) (https://ictv.global/ictv/proposals/2020.006D.R.Naldaviricetes.zip) (Walker et al., 2021). This class contains four families of nuclear arthropod large DNA viruses (NALDVs) including *Baculoviridae, Nadivirusidae, Hytrosaviridae*, and *Nimaviridae*, with AmFV as a free member. To place AmFV among the viruses, we conducted phylogeny analysis using concatenated protein sequences of the PIFs. The results showed that AmFV CN and AmFV CH-C05 formed a separate branch within *Naldaviricetes* (Fig. 3). Unlike the members of *Baculoviridae, Nadivirusidae, and Hytrosaviridae*, which also encode four subunits of a DNA-directed RNA polymerase, the AmFV CN and CH-C05 genomes lack DNA polymerase encoding these gene expression factors and thus, do not belong to the proposed order *Lefavirales* (https://ictv.global/ictv/proposals/2020.006D.R.Naldaviricetes.zip) (Walker et al., 2021). The PIF phylogenetic tree shows that AmFVs are distinct from the members of *Lefavirales*, although the AmFV PIFs appeared to be closer to those of *Hytrosaviridae* with a bootstrap value of 69% (Fig. 3).

### 3.4. The AmFV CN proteome

Since most of the ORFs of AmFV remain hypothetical, proteomics was performed to determine which ORFs are authentic proteins and whether they are structurally related. The proteins consistently identified in independent LC-MS/MS experiments are summarized in Table 3; 47 proteins were found to be associated with the AmFV virion (Table 3, Fig. 2). All the annotated PIFs, PIF0 (AmFV_077), PIF1 (AmFV_060), PIF2 (AmFV_100), PIF3 (AmFV_088), PIF4 (AmFV_157), and PIF5 (AmFV_057) were detected by proteomics (Table 3), suggesting that they are virion structural proteins that likely function as entry factors. AmFV_023, encoding a putative AAA + ATPase, was also identified (Tables 2 and 3). The functions of the remaining detected proteins are largely unknown (Table 3).
| Putative function | ORF size (aa) | Putative protein | Best match with Pfam-A database | Best match with BLASTP | AmFV CH-C05 identity |
|------------------|--------------|-----------------|---------------------------------|------------------------|----------------------|
| **DNA**          |              |                 |                                 |                        |                      |
| AmFV_027 610     | Thymidylate synthase | Thymidylat_synth | 1.5E-85  | PF00303.19 | Malassezia pachydermatis 284 3.0E-84 45.5% XP_017990541.1 98.2% |
| AmFV_042 556     | Integrase     | phage_integrase | 2.5E-05  | PF00589.22 | Fibrobacter sp. 55.1 2.0E-04 27.2% NLD99342.1 99.1% |
| AmFV_074         |              | DNA Pol         | 4.5E-14  | PF00361.21 | Athalia rose 99 9.0E-17 24.0% XP_025602669.1 99.6% |
| AmFV_095 1603    | DNA ligase    | DNA_ligase_A_N | 8.3E-09  | PF00467.14 | Cytophagaceae bacterium 730 0.0E-00 44.4% ODS80019.1 99.7% |
| **Replication**  |              |                 |                                 |                        |                      |
| and nucleotide   |              |                 |                                 |                        |                      |
| AmFV_114 879     | RR1           | Ribonuc_red_igC | 0.0E-00  | PF00867.15 | Uncultured virus 291 2.0E-121 100.0% AD576390.1 100.0% |
| **PIFs**         |              |                 |                                 |                        |                      |
| AmFV_057 334     | PIF-5         | –               | –                                 | –                      | Tipula deraeae nudivirus 170 4.8E-06 20.0% YP_009116743.1 99.1% |
| AmFV_060 830     | PIF-1         | PIF             | 7.0E-16  | PF005092.12 | Cyclophaga undans nucleopolyhedrovirus 91 4.0E-15 30.0% YP_010086341.1 98.6% |
| **BROs**         |              |                 |                                 |                        |                      |
| AmFV_008 182     | BRO-1         | –               | –                                 | –                      | Chrysoideixia chalcites nucleopolyhedrovirus 53 1.0E-04 33.3% YP_249718.1 100.0% |
| AmFV_016 1306    | BRO-2         | Bro-N           | 2.9E-09  | PF02498.17 | Helicoverpa armigera nucleopolyhedrovirus 179 1.60E-6 29.1% AMN15974.2 94.0% |
| **Others**       |              |                 |                                 |                        |                      |
| AmFV_069 262     | BRO-3         | –               | –                                 | –                      | AmFV 1309 0.0E-00 98.9% YP_009165820.1 98.9% |
| AmFV_075 434     | BRO-4         | –               | –                                 | –                      | Spodoptera littoralis nudivirus 52 9.0E-05 28.0% YP_001257066.1 98.9% |
| AmFV_106 667     | BRO-5         | Bro-N           | 3.2E-07  | PF02498.17 | Chrysoideixia inclusdens nucleopolyhedrovirus 77 3.0E-12 25.0% AGL57177.1 95.8% |
| AmFV_108 627     | BRO-6         | Bro-N           | 8.6E-04  | PF02498.17 | Spodoptera frugiperda ascovirus 1a 66 6.0E-09 29.0% YP_762434.1 92.3% |
| AmFV_110 499     | BRO-7         | Bro-N           | 6.6E-29  | PF02498.17 | Chrysoideixia chalcites nucleopolyhedrovirus 87 2.0E-15 29.8% AGE16781.1 97.0% |
| AmFV_111 437     | BRO-8         | –               | –                                 | –                      | AmFV 201 1.9E-14 25.2% YP_009165857.1 98.4% |
| AmFV_133 157     | BRO-9         | Bro-N           | 4.1E-03  | PF02498.17 | Acanthamoeba polyphaga mimivirus 111 6.7E-03 26.0% AKI80069 95.0% |
| AmFV_006 1699    | Protein kinase (PK) | –         | –                                 | –                      | Parasemium bursaria Chlorella virus CrI 179 2.4E-06 29.0% AGE16781.1 97.0% |
| AmFV_043 1633    | Myristoylated membrane | –          | –                                 | –                      | Mimivirus sp. SH 186 6.2E-12 46.7% AZL9416.1 96.0% |
| AmFV_089 572     | RING finger protein 413R | zf-CSHC4_3 | 2.1E-03  | PF13920.6 | Colliclithys fasciata 52 1.0E-03 33.3% TKS625457.1 98.6% |
| AmFV_082 510     | RING finger protein | –               | –                                 | –                      | Orcytne rhinoceros nudivirus 85 3.0E-15 31.1% YP_002321369.1 99.8% |
| AmFV_101 1982    | hypothetical protein | Gamma-glutamyltranspeptidase | 1.6E-14 | PF01019.21 | Hirsund rustica rustica 144 5.0E-04 40.2% RMB88007.1 99.8% |
| AmFV_113 583     | hypothetical protein | –               | –                                 | –                      | Diachasmimorpha longicaudata entomopoxovirus 147 4.5E-07 34.6% AKS26328.1 97.5% |
| AmFV_123 948     | hypothetical protein | –               | –                                 | –                      | Thalassosira oceanica 145 3.5E-08 32.9% EJK57244.1 90.5% |
| AmFV_168 1236    | MySHV 070     | –               | –                                 | –                      | Harpegnathus saltator 122 1.0E-04 29.4% EFN83926.1 99.8% |
| AmFV_193 969     | Chitin-binding | LOMP_10         | 8.6E-36  | PF003067.15 | Musca hytrosavirus 166 2.3E-04 29.6% YP_001883398.1 97.3% |
| AmFV_235 334     | PLC           | PI-PLC-X        | 3.6E-11  | PF00588.19 | Aipta mellifera 136 2.0E-31 42.4% WP_180560000.1 96.0% |
Further more, we detected 14 proteins with predicted molecular weights of 12.6–148.7 kDa with over 50% sequence coverage in at least one of the LC-MS/MS runs: AmFV_021 (36.8 kDa), AmFV_022 (30.6 kDa), AmFV_023 (69.2 kDa), AmFV_051 (39.1 kDa), PIF5 (AmFV_057, 36.1 kDa), AmFV_058 (148.7 kDa), AmFV_099 (42.6 kDa), AmFV_128 (56.6 kDa), AmFV_130 (19.8 kDa), AmFV_138 (53.4 kDa), AmFV_146 (24.9 kDa), AmFV_149 (12.6 kDa), AmFV_151 (44.0 kDa), and AmFV_156 (29.3 kDa), suggesting they are likely major structural proteins (Table 3).

4. Discussion

AmFV is an endemic DNA virus in honeybee colonies; however, very little was known about its molecular biology. In China, the prevalence of AmFV in honeybee colonies varies from 10% to 85% (Hou et al., 2016, 2017). In this study, we have produced useful information on this mysterious, large DNA virus.

The overall high nucleotide identity (96.9%) between the CN and CH–C05 strains indicates that these two viruses are closely related. Thirty-two of the 45 ORFs missing in AmFV CN (Table 1) encoded ORFs of less than 100 residues, suggesting that they may not be authentic ORFs. The gene order of AmFV CN was similar to that of AmFV CH–C05. However, unlike circular genome reported for the CH–C05 strain (Gauthier et al., 2015), our sequence data suggested that AmFV CN has a linear genome, as no reads overlapped both ends. In fact, our data support an earlier hypothesis that AmFV contains a linear genome based on electron microscopic observations of AmFV DNA (Bailey et al., 1981). Given that the length of the nucleocapsid is about 3.1 μm (Clark, 1978; Bailey et al., 1981; Sitaropoulouab et al., 1989) and the measured DNA length is 5.8 ± 0.3 μm (Bailey et al., 1981), viral DNA appeared to be only slightly condensed during packaging. How AmFV DNA interacts with proteins to form nucleocapsids, and how the long nucleocapsid is wrapped in the virion envelope into three figure-eight loops, remain open and intriguing questions.

Identifying structural proteins is fundamental for understanding the unique structure of AmFV and will provide useful information for the development of immunological AmFV diagnostic kits. By proteomic analyses, we identified 47 structural proteins that were present in at least two replicate runs (Table 3). As expected, all six PIFs (PIF 0–5) were present, suggesting that they are AmFV structural proteins and likely functional entry factors. In addition, AmFV_023, with homology to the ATPases associated with diverse cellular activities (AAA-) ATPase family, was identified. These ATPases are a large protein family that utilize energy from ATP hydrolysis to participate in multiple cellular functions. The presence of AmFV_023 in the AmFV virion suggests that it may be involved in providing required energy for nucleocapsid packaging.

Among the 47 proteins, 14 with predicted molecular weight of 12.6–148.7 kDa appeared to be major structural proteins, as they were
detected with over 50% sequence coverage (Table 3). Previous study has shown that 12 proteins, ranging in size from 13 to 70 kDa, had been associated with the AmFV virion by polyacrylamide gel electrophoresis (Bailey et al., 1981). Of these 12, P40 and P13 appeared to be major nucleocapsid proteins (although P13 was also present in the envelope), while P37 and P23 appeared to be major envelope proteins (Bailey et al., 1981). We speculate that AmFV_149 (predicted 12.7 kDa), AmFV_146 (24.9 kDa), PIF5 (36.1 kDa), and AmFV_51 (39.1 kDa) reflect the previously identified P13, P23, P37, and P40, respectively. However, this speculation is based solely on the rough matches in molecular weight and need verification in the future. One way to do this is to generate specific antibodies for probing western blots or performing immunoelectron microscopy.

In summary, our study showed that AmFV CN contains a linear genome of ~492,752 bp, encoding 197 ORFs. Forty-seven of the ORFs were associated with virions, including six PIFs, which likely function in viral entry. The results provide fundamental information for future molecular studies on the virus.

### Data availability

The genome sequence of AmFV CN has been deposited in GenBank under accession number OK392616.

### Ethics statement

This article does not contain any studies with human or animal subjects (except insects) performed by any of the authors.

### Author contributions

Dahe Yang: investigation, data curation, conceptualization, formal analysis, writing-original draft, writing-review and editing. Jun Wang: data curation, formal analysis, investigation, writing-original draft, writing-review and editing. Xi Wang: investigation, methodology. Fei Deng: funding acquisition, resources, supervision. Qingyun Diao: funding acquisition, resources, supervision. Manli Wang: funding acquisition.

| No. | ORFs* | Protein | Mol. weight (kDa) | Protein length (aa) | Test 1 | Unique peptides | Sequence coverage (%) | Test 2 | Unique peptides | Sequence coverage (%) |
|-----|-------|---------|------------------|--------------------|--------|----------------|----------------------|--------|----------------|----------------------|
| 1   | AmFV_002 | hypothetical protein | 108.0 | 961 | 8 | 7.8 | 5 | 5.1 |
| 2   | AmFV_017* | hypothetical protein | 46.6 | 409 | 10 | 27.7 | 5 | 14.5 |
| 3   | AmFV_021* | hypothetical protein | 36.8 | 321 | 18 | 66.9 | 11 | 31.6 |
| 4   | AmFV_022* | hypothetical protein | 30.6 | 271 | 12 | 52.2 | 7 | 26.7 |
| 5   | AmFV_023* | AAA + ATPase | 69.2 | 648 | 35 | 57.1 | 29 | 48.2 |
| 6   | AmFV_051* | hypothetical protein | 39.1 | 342 | 18 | 52.5 | 14 | 45.2 |

* The ORFs with sequence coverage over 50% in at least one test were marked with *.
resources, supervision. Zhihong Hu: conceptualization, formal analysis, funding acquisition, resources, supervision, writing-original draft, writing-review and editing. Chunsheng Hou: conceptualization, funding acquisition, resources, supervision, writing-review and editing.

Conflict of interest
The authors declare that they have no conflict of interest.

Acknowledgements
This work was supported by the Open Research Fund Program of the State Key Laboratory of Virology of China (Grant No. 2019IOV004), the key Research Program of Frontier Sciences of Chinese Academy of Sciences (Grant No. QYZDJ-SSW-SMC021), and the National Natural Science Foundation of China (Grant No. 31900154 and 31572471). We are grateful to Pei Zhang of the Core Facility and Technical Support, Wuhan Institute of Virology for her technical support in transmission electron microscopy.

References
Abou Kubaab, R.M.G., Solaiman Khaled, B., Daher-Hjaij, N., Heinoun, K., Saponari, M., 2016. First detection of black queen cell virus, varroa destructor macula-like virus, Apis mellifera filamentous virus and noose ceraeae in syrian honey bees Apis mellifera syriaca. Bull. Insectol. 71, 217–224.
Bailey, L., 1982. Viruses of honeybees. Bee World 63, 165-173.
Bailey, L., Carpenter, J.M., Woods, R.D., 1981. Properties of a filamentous virus of the honey bee (Apis mellifera). Virology 114, 1-7.
Beaurepaire, A., Pirot, N., Doublet, V., Antunez, K., Campbell, E., Chantawannakul, P., Chejanovsky, N., Gajda, A., Heerman, M., Panziera, D., Smagghe, G., Yanee, O., de Miranda, J.R., Dalmon, A., 2020. Diversity and global distribution of viruses of the western honey bee, Apis mellifera. Insects 11, 239.
Bezzer, A., These, G., Vavory, F., Gaillard, J., Poulain, J., Drezen, J.M., Hermiot, E.A., 2015. The genome of the nucleopolyhedrosis-causing virus from tipula oleracea sheds new light on the multifamily. J. Virol. 89, 3008–3025.
Bideshi, D.K., Renault, S., Stasiak, K., Federici, B.A., Bigot, Y., 2003. Phylogenetic analysis approaches to managing honeybee disease. Nat Ecol Evol 1, 1250–1262.
Chejanovsky, N., Gajda, A., Heerman, M., Panziera, D., Smagghe, G., Yanee, O., de Miranda, J.R., Dalmon, A., 2020. Diversity and global distribution of viruses of the western honey bee, Apis mellifera. Insects 11, 239.
Kuzio, J., Pearson, M.N., Harwood, S.H., Funk, C.J., Evans, J.T., Slavicek, J.M., Rohlmann, G.F., 1999. Sequence and analysis of the genome of a baculovirus pathogenic for lymantria dispar. Virology 253, 17–34.
Laughton, A.M., Siva-Jothy, M.T., 2011. A standardized protocol for measuring phenoloxidase and prophenoloxidase in the honey bee, Apis mellifera. Apidologie 42, 140–149.
Li, Y., Liu, X., Tang, P., Zhang, H., Qin, Q., Zhang, Z., 2021. Genome sequence and organization of the mythimna (formerly pseudalina) unipartita granulovirus Hawaiian strain. Sci. Rep. 11, 414.
Lietze, V.U., Abd-Alla, A.M., Vreysen, M.J., Geden, C.J., Bociais, D.G., 2011. Salivary gland hypertrophy viruses: a novel group of insect pathogenic viruses. Annu. Rev. Entomol. 56, 63–80.
McMenamin, A.J., Genersch, E., 2015. Honey bee colony losses and associated viruses. Curr Opin Insect Sci 8, 121–129.
Prodelalova, J., Moutelikova, R., Titera, D., 2019. Multiple virus infections in western honeybee (Apis mellifera L.) ejaculate used for instrumental insemination. Viruses 11, 306.
Quintana, S.B.C., Porrim, L.P., Di Geréimio, V., Eguarasar, M.J., Maggi, M., 2019. First molecular detection of Apis mellifera filamentous virus (amfv) in honey bees (Apis mellifera) in Argentina. J. Apicult. Res. 58, 211–217.
Ravonet, J., De Smet, L., Meneus, L., Smagghe, G., Wenseleers, T., de Graaf, D.C., 2014. Widespread occurrence of honey bee pathogens in solitary bees. J. Invertebr. Pathol. 122, 55–58.
Remnant, E.J., Shi, M., Buchmann, G., Blacquiere, T., Holmes, E.C., Beekman, M., Lietze, V.U., Abd-Alla, A.M., Vreysen, M.J., Geden, C.J., Boucias, D.G., 2011. Salivary gland hypertrophy viruses: a novel group of insect pathogenic viruses. Annu. Rev. Entomol. 56, 63–80.
McMenamin, A.J., Genersch, E., 2015. Honey bee colony losses and associated viruses. Curr Opin Insect Sci 8, 121–129.
Prodelalova, J., Moutelikova, R., Titera, D., 2019. Multiple virus infections in western honeybee (Apis mellifera L.) ejaculate used for instrumental insemination. Viruses 11, 306.
Quintana, S.B.C., Porrim, L.P., Di Geréimio, V., Eguarasar, M.J., Maggi, M., 2019. First molecular detection of Apis mellifera filamentous virus (amfv) in honey bees (Apis mellifera) in Argentina. J. Apicult. Res. 58, 211–217.
Ravonet, J., De Smet, L., Meneus, L., Smagghe, G., Wenseleers, T., de Graaf, D.C., 2014. Widespread occurrence of honey bee pathogens in solitary bees. J. Invertebr. Pathol. 122, 55–58.
Remnant, E.J., Shi, M., Buchmann, G., Blacquiere, T., Holmes, E.C., Beekman, M., Lietze, V.U., Abd-Alla, A.M., Vreysen, M.J., Geden, C.J., Boucias, D.G., 2011. Salivary gland hypertrophy viruses: a novel group of insect pathogenic viruses. Annu. Rev. Entomol. 56, 63–80.
McMenamin, A.J., Genersch, E., 2015. Honey bee colony losses and associated viruses. Curr Opin Insect Sci 8, 121–129.
Prodelalova, J., Moutelikova, R., Titera, D., 2019. Multiple virus infections in western honeybee (Apis mellifera L.) ejaculate used for instrumental insemination. Viruses 11, 306.
Quintana, S.B.C., Porrim, L.P., Di Geréimio, V., Eguarasar, M.J., Maggi, M., 2019. First molecular detection of Apis mellifera filamentous virus (amfv) in honey bees (Apis mellifera) in Argentina. J. Apicult. Res. 58, 211–217.
Ravonet, J., De Smet, L., Meneus, L., Smagghe, G., Wenseleers, T., de Graaf, D.C., 2014. Widespread occurrence of honey bee pathogens in solitary bees. J. Invertebr. Pathol. 122, 55–58.
Remnant, E.J., Shi, M., Buchmann, G., Blacquiere, T., Holmes, E.C., Beekman, M., Aso, A., 2017. A diverse range of novel rna viruses in geographically distinct honey bee populations. J. Virol. 91, e01158–17.
Sitrapeuleab, N., Neophytouab, E.P., Thomopoulosab, G.N., 1989. Structure of the nucleocapsid of a filamentous virus of the honey bee (Apis mellifera). J. Invertebr. Pathol. 53, 354–357.
Walker, P.J., Siddell, S.G., Letkowitz, E.J., Mushegan, A.R., Adriaienssens, E.M., Alifens-Zerbini, P., Davison, A.J., Dempsey, D.M., Dutlib, B.E., Garcia, M.L., Harrach, B., Harrison, R.L., Hendrickson, R.G., Junglen, S., Knowles, N.J., Krupovic, M., Kuhn, J.H., Lambert, A.J., Lobocha, M., NIHBL, M.L., Oksanen, H.M., Orton, R.J., Robertson, D.L., Rubino, L., Saganazovick, S., Simmonds, P., Smith, D.B., Suzuki, N., Van Doerslaer, K., Vandame, A.M., Vreysen, M.J., Azeni, F.M., 2021. Changes to virus taxonomy and to the international code of virus classification and nomenclature ratified by the international committee on taxonomy of viruses (2021). Arch. Virol. 166, 2633–2648.
Wang, X.-Z., Shang, X., Chen, C., Liu, S., Zhang, M., Zhang, N., Hu, H., Zhang, F., Zhang, T., Wang, Z., Liu, X., Lin, Z., Deng, F., Wang, H., Zou, Z., Vlak, J.M., Wang, M., Hu, Z., 2019. Baculovirus per os infectivity factor complex: components and assembly. J. Virol. 93, e02053–18.
Wang, Y., Jehle, J.A., 2009. Nudiviruses and other large, double-stranded circular DNA viruses of invertebrates: new insights on an old topic. J. Invertebr. Pathol. 101, 187–193.
Zemskov, E.A., Kang, W., Maeda, S., 2000. Evidence for nucleic acid binding ability and nucleosome association of bombus mori nucleopolyhedrovirus boj proteins. J. Virol. 74, 6784–6789.