CryoSeekV: Discovery of a novel virus AaTLV-IMCAS from *Aedes albopictus* cells by cryogenic electron microscopy

Hao Qu[a,b,1], Linjie Li[a,1], Xiaoming Li[a,c], Yunhan Xi[a,b], Ruirui Yang[a,d], Xiao Qu[a], Qiuyao Jiang [e], Peiyi Wang [f],Shihua Li[a,*] and George F. Gao[a,g,*]

[a] CAS Key Laboratory of Pathogen Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences (CAS), Beijing 100101, China;
[b] Medical School, University of Chinese Academy of Sciences (UCAS), Beijing 100101, China;
[c] Department of Gastroenterology, Shenzhen Children's Hospital, Shenzhen 518000, China;
[d] Shanxi Key Laboratory of Animal Disease Research, Prevention and Control, College of Veterinary Medicine, Shanxi Agricultural University, Jinzhong 030801, China;
[e] Medical Science and Technology Innovation Center, Shandong First Medical University & Shandong Academy of Medical Sciences, 250000 Jinan, China;
[f] Institute of Crop Microstructure Research, Shandong Agricultural University, Tian 271018, China;
[g] Institute of Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention (China CDC), Beijing 102206, China;
[1]H.Q. and L.L. contributed equally to this work.

Author Contributions:

H.Q., S.L., and G.F.G. designed the research; H.Q., L.L., and S.L. performed the research; Y.X., R.Y., and X.Q. contributed new reagents/analytic tools; H.Q., L.L., X.L., and Y.X. analyzed data; H.Q., L.L., and S.L. performed experimental validation; and H.Q., S.L., and G.F.G. wrote and revised the paper.

Correspondence
George F. Gao, CAS Key Laboratory of Pathogen Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, 100101 Beijing, China.
[*]**To whom correspondence may be addressed.** Email: lish@im.ac.cn or gaof@im.ac.cn.

Competing Interest Statement: The authors declare no competing interests.

Classification: Biological Sciences; Microbiology.

Keywords: Virus discovery, Cryo-EM, ModelAngelo, CryoSeekV, Structural biology

Abstract

Discovering divergent RNA viruses remains challenging due to the limitations of reference-based sequencing and viral isolation. Here, we present the structure-guided discovery of *Aedes albopictus*-associated *Tymoviridae*-like virus strain IMCAS (AaTLV-IMCAS), isolated from C6/36 mosquito cell supernatant. Without prior genomic information, we determined its capsid structure at a resolution of 2.11 Å using cryogenic electron microscopy (cryo-EM). Deep-learning-based automated modeling via ModelAngelo generated position-specific residue profiles that directed the targeted capture and assembly of the complete 6,488-nucleotide positive-sense single-stranded RNA genome, establishing a structural discovery workflow named CryoSeekV. Phylogenetic analysis showed that AaTLV-IMCAS, along with the related mosquito-borne viruses, forms an independent monophyletic clade distinct from established plant-infecting genera, representing a novel genus; therefore, we proposed the genus *Mosquivirus*, within the family *Tymoviridae*. Structural analysis revealed a $T = 3$ icosahedral capsid where a 9 Å main-chain shift within the N-terminus acts as a conformational switch governing quasi-equivalent capsomer assembly. Functional assays demonstrated a strict mosquito cell tropism and temperature-dependent replication optimized at 28 °C that establishes persistent infection without cytopathic effects. Time-resolved transcriptomics revealed that infection triggers dynamic host reprogramming: an early (24 h) surge in host translation, oxidative phosphorylation, and amino-acid metabolism coupled with selective repression of key melanization and recognition factors, which largely returned toward baseline by 48 h. Together, this study

establishes CryoSeekV as an effective paradigm for structure-driven viral surveillance and provides foundational molecular and physiological insights into invertebrate-associated *Tymovirales*.

Significance Statement

Traditional virus discovery relies heavily on genetic sequencing or virus isolation, which often fail when identifying divergent viruses lacking reference sequences. Here, we present the first structure-driven discovery and near-atomic characterization of a viral lineage directly from cellular reservoirs without prior sequence data. Combining cryogenic electron microscopy (cryo-EM) with artificial intelligence (AI) assisted modeling, we inverted the conventional virology workflow, demonstrating that high-resolution capsid structure can serve as the direct entry point to reverse-engineer primary sequences and capture complete viral genomes. This framework, named CryoSeekV here, uncovers a distinct insect-associated viral lineage and elucidates the molecular mechanics driving its capsid assembly. These findings establish a validated paradigm for structural viral surveillance and expand our understanding of viral dark matter, and further expand the cryo-EM application for the discovery of previously-uncharacterized bioentities.

Main Text

Introduction

The rapid expansion of global metagenomic surveillance with next generation sequencing (NGS) technology has revealed that the biosphere harbors an immense reservoir of uncharacterized RNA viruses, frequently referred to as "viral dark matter" (1-4). Within the +ssRNA viral kingdom *Orthornavirae*, numerous lineages possess genomic architectures and replication strategies that remain virtually unknown (5-7).

The discovery and classification of these divergent agents rely heavily on either conventional virus isolation or high-throughput NGS sequencing coupled with homology-based alignment against public reference databases (8). However, this traditional reliance on sequence-first or isolation-based workflows faces inherent limits with highly divergent viruses. Lacking detectable homology to reference families with vast NGS sequencing data, such genomes frequently suffer from assembly failures, reference bias and inaccurate annotation (9, 10). Furthermore, many latent or endogenous viruses exist at low initial concentrations within dense host genetic backgrounds, making their selective amplification and subsequent bioinformatic identification extremely challenging (11). For virus isolation it is also a tedious, sometimes unachievable task due to unavailable cell lines. Consequently, expanding our current understanding of the virosphere requires complementary methodologies capable of identifying and classifying divergent viral lineages independent of initial sequence alignment or virus isolations.

Over the past decade, single-particle cryogenic electron microscopy (cryo-EM) has undergone a resolution revolution (12-15), driven by the development of direct electron detectors (DED) and robust image processing platforms such as RELION (16) and cryoSPARC (17). Unlike traditional X-ray crystallography, cryo-EM allows for the direct structural visualization of macromolecular complexes and viral particles in a near-native state without the prerequisite of crystallization. In structural virology, cryo-EM has become an indispensable tool for resolving the atomic architectures of viral capsids, envelope glycoproteins and receptor-bound entry intermediates (7).

107 Importantly, the implementation of 2D and 3D classification algorithms provides an
108 inherent advantage termed *in silico* purification, allowing researchers to
109 computationally segregate heterogeneous populations or low-abundance structural
110 entities from complex, non-purified biological samples. Consequently, the capability to
111 routinely resolve electrostatic potential maps at near-atomic or atomic resolutions (2–3
112 Å) has laid the foundation for extracting primary chemical information directly from
113 three-dimensional structural data.

114 This continuous improvement in map resolution has driven the rapid advancement
115 of computational technologies designed for sequence recognition directly from electron
116 density maps. At near-atomic thresholds, the spatial distribution of electrostatic
117 potentials can resolve detailed side-chain features, enabling the direct mapping of
118 amino acid identities. Early software frameworks, exemplified by automated modeling
119 suites within PHENIX (18), relied heavily on backbone tracing followed by manual
120 side-chain fitting under the constraint of a known target sequence. To overcome the
121 limitation of requiring a predetermined reference sequence, database-matching
122 algorithms such as cryoID (19) and findMySequence (20) were developed to extract
123 short peptide tags or hidden Markov model (HMM) profiles from density characteristics,
124 thereby identifying unknown components within macromolecular complexes. Recently,
125 this technical trajectory achieved a major milestone with the introduction of deep-
126 learning-based automated modeling frameworks. Tools such as ModelAngelo (21)
127 utilize graph neural networks (GNNs) to combine local density features with global

spatial geometry, predicting position-specific amino acid probabilities directly from unassigned density maps and opening new avenues for *de novo* sequence determination.

The convergence of high-resolution cryo-EM and deep-learning sequence recognition has recently enabled a conceptually novel paradigm: structure-driven bioentity discovery. This concept was robustly demonstrated by the development of the CryoSeek strategy (22, 23), which proved that uncharacterized proteins within complex environmental water samples could be modeled *de novo*, reverse-sequenced via structural alignment databases like Foldseek (24), and successfully identified independent of prior genomic data. While traditional structural biology typically follows a functional or sequence-driven path ("sequence-structure-function"), structure-driven discovery reverses this workflow into a "structure-sequence-biology" pipeline. Despite its clear potential for uncovering biological dark matter, the systematic application of this structure-guided approach to virus discovery specifically for capturing divergent viral genomes and detailing their underlying biology remains largely unexplored.

The family *Tymoviridae* (order *Tymovirales*) traditionally comprises +ssRNA viruses that predominantly infect plant hosts and rely on insect vectors for mechanical transmission. These viruses characteristically possess a genome of 6.0–7.5 kb encoding a replication polyprotein and a single coat protein that assembles into a triangulation number of 3 ($T = 3$) icosahedral shell composed of 180 subunits. Although recent sequencing-based ecological surveillance has hinted at the presence of "insect-related *Tymoviridae*-like viruses", such as *Culex Tymo-like virus strain Los Angeles* (CTymVL)

(25), *Culex Tymo-like virus* (CuTLV) (26), *Ek Balam virus* (EkBV) (27), and *Guachaca virus* (GUAV) (28), these lineages completely lack high-resolution structural characterization. This structural void creates a substantial gap in the field: without atomic-resolution details, it is difficult to resolve the evolutionary mechanisms of how these viruses transitioned from plant-associated frameworks to establish persistent infections within insect cell lines, or how their single coat proteins undergo localized structural changes to accommodate the sharp geometric constraints of pentameric versus hexameric capsomers.

In this study, we applied a structure-guided discovery pipeline named CryoSeekV to identify an uncharacterized, persistent viral agent uncovered in C6/36 mosquito cell cultures, designated *Aedes albopictus*-associated *Tymoviridae*-like virus strain IMCAS (AaTLV-IMCAS) (Fig. 1). Without utilizing initial sequence data, we resolved the capsid structure of AaTLV-IMCAS at a resolution of 2.11 Å and used ModelAngelo to derive an accurate structural sequence profile directly from the density map. This structural information guided reverse transcription-PCR (RT-PCR) and targeted NGS strategies to obtain the complete 6,488-nucleotide genome. Evolutionary and genomic analyses demonstrated that AaTLV-IMCAS, together with related mosquito-borne viruses, forms a highly distinct clade that branches separately from all established plant-infecting *Tymovirale* genera, representing a potential novel genus, which we propose to name genus *Mosquivirus* within Family *Tymoviridae*. Furthermore, our structural and functional analyses revealed the precise molecular interactions supporting its $T = 3$ capsid assembly, characterized a critical 9 Å N-terminal conformational switch, and characterized its strict insect-specific biological tropism as well as the dynamic expression profile following host infection. This work establishes a validated paradigm for structure-driven viral surveillance and provides foundational insights into the evolution of invertebrate-associated *Tymoviridae.*

Results

Structural Delimitation of Capsid Symmetry and Structure-Driven Sequence Identification

The uncharacterized virus was serendipitously uncovered during the amplification and purification of other flaviviruses using C6/36 cells and was found to replicate stably across successive passages. To determine its molecular identity, culture supernatants were collected for cryo-EM sample preparation. Following particle picking, multiple rounds of reference-free 2D classification and 3D heterogeneous refinement successfully segregated the viral particles from background noise, ultimately yielding a high-resolution electrostatic potential map (Fig. 1 and Fig. S1).

To characterize the molecular identity of the uncharacterized viral agent, we first analyzed the global topological features of the obtained 2.11 Å high-resolution electrostatic potential map. The viral particle possesses a classic triangulation number of 3 (T=3) icosahedral symmetry, with an outer shell diameter measuring approximately 350 Å (Fig. 2*A*). The continuous protein capsid shell is structurally organized by 180 structural subunits clustered into 12 pentameric capsomers situated at the five-fold vertices and 20 hexameric capsomers forming the face centers (Fig. 2*A*, 2B). Each standard asymmetric unit (ASU) is structurally defined as a rigid heterotrimer composed of three chemically identical coat protein (CP) subunits—designated as the pentameric vertex-associated core, the hexameric face-associated core-1, and the hexameric face-associated core-2—which adopt slightly different conformations to accommodate the localized geometric constraints (Fig. 2*B*). Clear secondary structure density features, characterized by continuous backbone connectivity and distinct side-chain protrusions, indicated that each monomeric subunit adopts a canonical jelly-roll β-barrel fold dominated by dense β-sheets. Intriguingly, while this structural assembly mode is common among plant viruses, it is rare among insect or animal viruses, creating a stark contrast to its ability to propagate in C6/36 cells.

To extract the primary sequence directly from the unassigned density map independently of prior genomic sequencing data, the electron density map of a single capsid monomer was segmented (Fig. 2*C*). We initially attempted a structural homology search by taking the automated atomic model built by ModelAngelo (yellow; Fig. 2*C* and *D*) and querying it within the Protein Data Bank (PDB) via Foldseek. However, due to a profound lack of high-resolution atomic coordinates for insect-associated *Tymoviridae*-like lineages in structural databases, this structural alignment via Foldseek was exclusively restricted to plant-infecting members within the family *Tymoviridae*, retrieving *Citrus sudden death-associated virus* as a close match and aligning to a structurally similar homolog (blue; PDB ID: 7SQY; Fig. 1 and Fig. 2*E*) (29). Because typical plant *Tymoviruses* are ecologically restricted to plant hosts and fail to replicate inside dipteran systems, this structural comparison highlighted a distinct host tropism conflict.

To bypass this database bias and leverage higher-specificity primary sequence features, the automated model built by ModelAngelo successfully predicted sequential peptide fragments, providing sufficient amino acid sequence profiles to execute a high-specificity protein-protein BLAST (BLASTP) query against the NCBI database. In contrast to the fold-based alignment, this sequence-based profile bypassed the structural database limitation, directly matching uncharacterized coat proteins of *Culex Tymo-like virus* (CuTLV)(26) and accurately mapping the agent to the insect-associated *Tymoviridae*-like virus lineage.

To validate this structure-driven sequence annotation and fully capture the viral genome, total RNA extracted from the purified viral particles was subjected to NGS, and raw transcriptomic reads were assembled using the automated structural sequence as a spatial template (Fig. 1). Gaps within low-coverage regions were resolved by targeted RT-PCR and continuous Sanger sequencing, completely validating the continuous 6,488-nucleotide genome. Global alignment against the validated primary sequence revealed that the AaTLV-IMCAS exhibits 82.69% nucleotide identity and 91.13% coat protein amino acid identity with the previously reported *Culex Tymoviridae-like virus*. Crucially, manual refinement of the atomic model using the validated primary sequence demonstrated that the initial automated structure predicted by ModelAngelo was remarkably close to the experimentally refined true model, characterized by an overall root-mean-square deviation (RMSD) of only 0.311 Å and an amino acid sequence match identity of 71.1% (Fig. 2*D* and *F*). This high sequence congruence and excellent stereochemical alignment powerfully illustrated that structure-driven sequence recognition possesses superior specificity and accuracy for the identification of divergent viral lineages.

Molecular Switch Mechanism and Thermodynamic Hierarchy Driving Quasi-Equivalent Capsid Assembly

To systematically elucidate how identical capsid protein monomers differentiate to form the T=3 ASU, we comprehensively analyzed the targeted spatial packaging and inter-subunit contact interfaces of the AaTLV-IMCAS virion. The heterotrimeric asymmetric unit is organized by three distinct chemical environments: the pentameric vertex-associated core, the hexameric face-associated core-1, and the hexameric face-associated core-2 (Fig. 3*A*). These three core components cluster dynamically through an intricate arrangement of localized bonds to establish a stable structural building block (Fig. 3A). Interface analysis mapped via the PDBePISA server identified three clear, non-equivalent binding junctions within this structural trimer (Fig. 3*B*, *C* and *D*). Specifically, the interface between the pentameric core and hexameric core-1 is tightly anchored by multiple hydrogen bonds formed between pairs S84–T78, N85–T78, P203–T207, and Y244–R195 (Fig. 3*B*). Similarly, the junction between the pentameric core and hexameric core-2 replicates a parallel docking profile, utilizing pairs T78–N85, R195–Y244, and T207–R203 to lock the polypeptide chains (Fig. 3*C*). In contrast, the contact zone directly bridging hexameric core-1 and hexameric core-2 features a simplified coordinating network primarily mediated by a singular hydrogen bond between R195 and T78 (Fig. 3*D*). This spatial distribution reveals a definitive thermodynamic stability gradient within the asymmetrical trimer block, decreasing from the rigid pentamer-hexamer contacts to the highly malleable hexamer-hexamer interface.

On a macroscopic scale, these asymmetric building blocks multimerize into two morphologically distinct classes of higher-order capsomers across the capsid shell. At the five-fold symmetry axes, five subunit copies assemble into a steep, pyramidal pentameric capsomer with an outer ridge dimension of approximately 57.6 Å (Fig. 3*E*). In contrast, six subunit copies cluster around the local quasi-three-fold axes into hexameric capsomers displaying quasi-six-fold symmetry and an outer ridge span of 53.9 Å (Fig. 3*F*). This pronounced pyramidal elevation allows the five subunits to tilt inward and pack more tightly around the vertex center, resulting in a more rigid and stable assembly for the pentameric capsomeres compared to the flatter hexamers.

To unravel the micro-chemical details supporting these structural variations, we systematically quantified the respective coordinating bonding residues across the isolated symmetric faces (Fig. 3*G*, *H*, and *I*). Within the pentamer, five subunits are structurally equivalent and display an identical, dense interaction matrix. A single pentameric monomer coordinates 13 distinct hydrogen bonds (< 3.5 Å; involving residues D66, T149, E151, M167, E168, S169, S174, G178, S185, S187, and L218 against neighboring strands D63, Q95, T97, R140, Q141, P180, I181, F183, and N186) and 1 stable salt bridge (anchored by the E151–R140 pair) to guarantee capsid rigidity (Fig. 3*G*). In contrast, the internal hexameric junctions reveal chemical heterogeneity and are partitioned into two distinct interaction configurations (Fig. 3*H* and *I*). The first hexameric interface (Fig. 3*H*) coordinates 8 hydrogen bonds (< 3.5 Å; involving F69, T149, E151, M167, E168, S174, S176, and L218 paired with M62, Q95, T97, P139, R140, and Q141) and 1 salt bridge (formed by pair E151–R140). The second hexameric interface (Fig. 3I) coordinates 11 hydrogen bonds (< 3.5 Å; involving M62, Q95, T97, K117, P139, R140, and Q141 paired with F69, T149, E151, M167, E168, S169, S176, and N219) and 2 salt bridges (formed by pairs K117–E168 and R140–E151). This localized reduction in bonding density confirms that the hexameric face blocks possess significantly lower chemical stability and higher structural plasticity than the pentameric vertices, which is consistent with our previous analysis.

To resolve the precise structural mechanics driving this capsomer differentiation, we aligned the atomic structures of the three independent monomeric subunits representing the different symmetric microenvironments. Although the core jelly-roll β-barrel platform remains completely identical, localized conformational variations are prominently restricted to the flexible loop regions at the N- and C-termini. At the extreme C-terminus, residues 246–248 flip upward exclusively in the pentameric subunit to alleviate localized spatial crowding, while remaining in a uniform downward conformation in both hexamer variants. More importantly, the flexible N-terminal arm (residues 55–89) undergoes major structural polymorphism. Owing to varying spatial confinements, the electron density of the flexible N-terminus remains structured as far as residue 55 in hexamer core-1 and residue P61 in hexamer core-2, but becomes completely disordered prior to residue D63 in the tightly packed pentamer environment.

Crucially, between residues 55 and 77, the polypeptide main chain of this N-terminal loop undergoes a dramatic 9 Å spatial displacement between the pentameric and hexameric configurations (Fig. 3J). Comprehensive residue mapping verified that the amino acid residues driving capsomer assembly (intra-pentamer and intra-hexamer) are virtually identical in the central domain (spanning residues 140–174), but diverge drastically at the N- and C-terminal margins (Fig. 3*K*). These data demonstrate that the N-terminal flexible arm functions as the definitive molecular switch that alternates localized subunit pairing paths, enabling a single coat protein sequence to successfully guide polymorphic shell assembly.

Genome Architecture, Domain Refinement and Evolutionary Allocation of a Potential Novel Genus

The AaTLV-IMCAS genome contains two major open reading frames (ORFs) (Fig. 4*A*). ORF1 (nucleotides 202–5,478) encodes a massive 1,759-amino-acid replication-associated polyprotein, while ORF2 (nucleotides 5,522–6,268) encodes the 249-amino-acid CP (Fig. 4*A*). RNA secondary structure profiling of the non-coding regions using the RNAfold server indicated that the 5' UTR (1–130 nt) forms a flexible stem-loop configuration with a minimum free energy (ΔG) of -10.80 kcal/mol (Fig. 4*B*), a hallmark of viral translation initiation elements. Conversely, the 3' UTR (6,269–6,488 nt) folds into a highly stable, complex stem-loop network ($\Delta G = -26.10$ kcal/mol) (Fig. 4*C*), likely serving as a rigid cis-acting replication element for viral polymerase recognition.

To precisely define the functional domain boundaries within the multi-functional ORF1 replication polyprotein, bioinformatics annotations were systematically reconciled with 3D structural models predicted by AlphaFold 3 (Fig. 4*D*-4*I*). This unified structural approach allowed for the refinement of domain borders based on structural fold continuity rather than primary sequence alignments alone. Specifically, the Methyltransferase domain was extended to span residues 38–339 due to the inclusion of a newly identified stable α-helix bundle (Fig. 4*E*). The boundaries of the Protease domain were refined to residues 658–791 following the structural omission of highly flexible, non-conserved terminal loops (Fig. 4*F*). The Helicase domain was accurately mapped to residues 807–1,149 based on the continuity of its core β-folding unit (Fig. 4*G*), and the catalytic RNA-dependent RNA polymerase (RdRp) domain was assigned to residues 1,206–1,733 to completely enclose its conserved fingers-palm-thumb catalytic core (Fig. 4*H*).

To determine the exact evolutionary trajectory of AaTLV, a maximum-likelihood phylogenetic tree was reconstructed under the GTR+G+I substitution model using the conserved RdRp amino acid sequences of representative members within the order *Tymovirales* (Fig. 4*J*). The phylogenetic topology demonstrated that AaTLV, together with three recently documented mosquito-borne *Tymoviridae-like* viruses (CuTLV, EkBV and GUAV), forms a tightly supported, monophyletic evolutionary cluster with a bootstrap value of 100% (Fig. 4*J*). Crucially, this insect-associated cluster (colored tan) branches completely separately from the three classic plant-infecting genera—*Tymovirus* (pink), *Maculavirus* (light purple), and *Marafivirus* (light green)—within the family *Tymoviridae*. This distinct branching profile provides robust evolutionary evidence that AaTLV and its related mosquito-borne viruses represent a distinct, independent lineage, strongly supporting its taxonomic assignment to a potential novel genus within the *Tymoviridae*-related evolutionary framework, named *Mosquivirus*.

Biological Characterization: Strict Host Tropism and Temperature-Dependent Persistent Infection

To investigate the biological phenotype and cellular pathogenesis of AaTLV-IMCAS, purified virions were inoculated onto native *Aedes albopictus* C6/36 mosquito cell monolayers. Over a continuous 72-hour monitoring period, mock-infected and virus-infected cells remained morphologically indistinguishable under light microscopy (Fig. 5A). No typical indicators of viral CPE—such as cell rounding, syncytia formation, detachment, or monolayer lysis—were observed at 24 h, 48 h, or 72 h post-infection (Fig. 5*A*). This complete lack of cytotoxicity demonstrates that AaTLV-IMCAS establishes a highly adapted, benign persistent infection within its insect host cells rather than undergoing a lytic replication cycle.

To map the host range and replication boundaries of the virus across different taxa, a highly sensitive qRT-PCR assay targeting a conserved 112-bp fragment within the RdRp domain was developed. Inoculation assays were conducted in parallel across a diverse panel, including C6/36 cells, feline kidney (CRFK) cells, baby hamster kidney (BHK-21) cells, African green monkey kidney (Vero) cells, and human embryonic kidney (293T) cells (Fig. 5*B*). Monitoring of viral growth kinetics demonstrated robust viral RNA amplification exclusively within the insect-derived C6/36 cell line, with intracellular viral RNA levels peaking between 60 and 72 hours post-infection (Fig. 5*B*). In contrast, all tested mammalian cell lines (BHK-21, 293T, CRFK and Vero) failed to support viral replication, with viral RNA copy numbers remaining flat or falling below the baseline detection threshold throughout the 72-hour course, confirming a strict insect-specific host tropism.

Furthermore, temperature shift experiments within C6/36 cells revealed that the replication efficiency of AaTLV-IMCAS is rigidly governed by environmental temperature constraints (Fig. 5*C*). The virus exhibited optimal, rapid replication kinetics exclusively at 28°C (Fig. 5*C*). However, raising the cultivation temperature to 33°C or 37°C completely abolished viral replication (Fig. 5*C*). Taken together, these functional data establish that AaTLV-IMCAS is an insect-restricted, temperature-restricted viral agent that is biologically optimized for persistent coexistence within mosquito cellular systems.

Time-Resolved Transcriptomic Profiling Reveals Transient Metabolic Activation and Immune Modulation Upon Viral Infection

We studied how *Aedes albopictus* C6/36 cells respond to viral infection using bulk RNA-sequencing of infected (Virus) and control (Mock) cultures at 24 and 48 h post-infection (two Mock and three Virus samples per timepoint). Principal component analysis clearly separated the samples and distinguished those four conditions well (Fig. 6*A*). Comparing Virus with Mock (adjusted $P < 0.05$ and $|\log 2 \text{ fold change}| > 1$), we found 267 differentially expressed genes (DEGs) at 24 h (154 up, 113 down) and 141 DEGs at 48 h (80 up, 61 down) (Fig. 6*B*, *C*). A total of 22 genes were shared between the two time points (Fig. 6*D*).

Several analyses confirmed this temporal split. On a t-SNE map, genes that changed only at 24 h, only at 48 h, or at both fell into separate regions (Fig. S2*A*); fold changes at the two timepoints were only weakly correlated (Fig. S2C); and the main blocks of the expression heatmap followed timepoint (Fig. S2B). Grouping genes by their 24 h and 48 h fold changes sorted them into simple temporal patterns — sustained, transient, late and reversed (Fig. S2*D*).

We then tested the 24 h DEGs for enriched functions. Up-regulated genes were linked to active metabolism and biosynthesis (amino-acid and carbon metabolism and L-serine pathways) and to innate-immune and melanin-biosynthesis terms (Fig. S3*A*, *B* and *D*). Ribosome, oxidative phosphorylation, and ER protein-processing gene sets were strongly induced, and several infection-related KEGG pathways were enriched (Fig. S3*C*). At the same time, key mosquito immune genes were down-regulated, including a phenoloxidase (a melanization enzyme; human ortholog *TYR*) and two fibrinogen-domain recognition proteins (human analogs *FGL1* and *MFAP4*).

A protein–protein interaction network of the 24 h DE proteins formed clear functional modules (oxidative phosphorylation, vesicle transport, amino-acid and fatty-acid metabolism, and cell cycle) built around hub proteins such as NDUFA1/4, PSMC4, VCP, RAB7A and HSP70 chaperones (human ortholog names are used for functional annotation) (Fig. 6*E*). Many pathways induced at 24 h had returned toward baseline by

48 h (Fig. 6*F*).

Finally, a two-factor model identified genes whose infection response changed between time points, such as the ER chaperone CALR (up at 24 h, back toward baseline at 48 h; Fig. S4*C*). The data were of good quality, with highly correlated replicates (Pearson $r \geq 0.97$; Fig. S4*A*) and well-behaved dispersion estimates (Fig. S4*B*), and the four DEG sets overlapped mostly within rather than across time points (Fig. S4*D*).

Discussion

The successful characterization of AaTLV-IMCAS demonstrates the efficacy of a structure-driven "structure – sequence – biology" framework for capturing and identifying uncharacterized viral lineages within complex biological systems. In classical virology, discovering low-abundance or endogenous viral entities embedded within dense host genetic backgrounds relies heavily on deep metagenomic sequencing (8). However, this conventional approach frequently encounters bottlenecks when sequences fall below alignment thresholds or when host transcripts outcompete viral genetic signals (11). By utilizing single-particle cryo-EM, the inherent advantage of *in silico* 2D or 3D classification allowed us to computationally segregate and purify the 35 nm AaTLV-IMCAS capsids directly from heterogeneous sample matrices, bypassing the rigorous requirements of empirical biochemistry. Combined with deep-learning automated tracing via ModelAngelo (21), the extraction of position-specific residue probabilities from a 2.11 Å potential map provided sufficient primary sequence tags to perform targeted transcriptomic retrieval. This CryoSeekV strategy underscores a crucial methodological shift: high-resolution structural virology can serve as a primary diagnostic tool to investigate the "viral dark matter" lurking as persistent contaminants in widely used laboratory cell lines.

Phylogenetically, the allocation of AaTLV within a highly supported, independent monophyletic clade alongside CuTLV(26), EkBV(27) and GUAV(28) strongly advocates for the establishment of a novel genus within the *Tymoviridae*-related evolutionary framework. Historically, the family *Tymoviridae* has been recognized almost exclusively as a group of plant-infecting pathogens that exploit specific insect vectors solely for mechanical or non-propagative transmission (30). The identification of a distinct, replication-competent lineage that branches completely separately from the established plant-infecting genera (*Tymovirus*, *Maculavirus* and *Marafivirus*) indicates a significant evolutionary divergence event, marking a transition from mere mechanical vectoring to active propagation within invertebrate hosts. This taxonomic grouping implies that ancestral *Tymoviruses* may have successfully transitioned from plant-associated ecological niches to achieve stable, permanent colonization of dipteran hosts by acquiring the machinery necessary to robustly replicate inside insect, and particularly mosquito, cells. Notably, while AaTLV-IMCAS is phylogenetically nested within this radiation, its capacity to cross kingdom boundaries and infect plant hosts remains entirely unverified, which is an open question shared by other mosquito-borne tymo-like counterparts (CTymVL, CuTLV, EkBV and GUAV). This evolutionary leap from non-propagative vectoring to autonomous replication strongly suggests that profound molecular adaptations must have occurred within both the replication-associated proteins and the host-receptor binding domains of the capsid. Future systematic investigation into these structural and functional variations will be imperative to fully unravel the molecular mechanisms driving this virus's cross-species transmission and host-tropism shift.

The biological characterization of AaTLV-IMCAS reveals a highly adapted, benign persistent relationship with its insect host. The complete absence of visible CPE in C6/36 cell monolayers over an extended 72-hour infection period is particularly remarkable given the genetic background of this specific cell line. The *Aedes albopictus* C6/36 clone is globally recognized to harbor a naturally occurring mutation within its *Dicer-2* gene, which abolishes the canonical exogenous siRNA pathway. The capability of AaTLV-IMCAS to accumulate to high viral titers within this RNAi-deficient system without triggering cytotoxicity indicates an advanced evolutionary equilibrium. This benign equilibrium is further substantiated by our time-resolved transcriptomic profiling. Rather than provoking sustained cytopathology, AaTLV-IMCAS infection elicits a transient transcriptional rewiring characterized by an early (24 h) surge in protein translation, oxidative phosphorylation, and amino-acid biosynthesis—likely hijacked to fuel efficient viral assembly—coupled with the down-regulation of key innate immune effectors such as phenoloxidase and fibrinogen-domain recognition proteins. Notably, this pronounced response largely attenuates toward baseline by 48 h, demonstrating that the host avoids prolonged stress responses and providing a molecular rationale for the non-lytic, persistent phenotype observed in vitro. Crucially, because C6/36 cells are universally deployed for the isolation and expansion of major medical arboviruses but traditionally lack robust, non-toxic genetic engineering tools, this non-lytic phenotype renders the AaTLV-IMCAS capsid an exceptionally attractive candidate for developing insect-specific viral vectors. Such tools could provide a practical and benign platform for targeted gene delivery, gene editing, or functional genetic screens exclusively within dipteran systems.

Beyond its potential utility as a bioengineering tool, the persistent presence of AaTLV-IMCAS raises critical questions regarding its impact on the cellular landscape during co-infection. Given its high-density and stealthy replication within C6/36 cells, it remains to be determined whether AaTLV-IMCAS infection induces superinfection exclusion or otherwise affects the replication kinetics and host immune response of heterologous, human-pathogenic arboviruses. Investigating the molecular cross-talk and potential viral interference between this virus and major public health threats (e.g., *Yellow fever virus*, *Dengue virus*) in co-infected systems will be a critical avenue to evaluate whether AaTLV-IMCAS alters vector susceptibility or serves as a natural modulator of arboviral transmission.

Finally, while this study establishes a robust paradigm CryoSeekV for structure-driven virus discovery, a fundamental gap remains between our *in vitro* observations and *in vivo* ecological realities. Our current functional assays and phenotype evaluations were restricted entirely to continuous cell lines. Consequently, whether AaTLV actively circulates within wild mosquito populations, as well as its precise tissue tropism, vector competence, and vertical or horizontal transmission dynamics within live, colonized *Aedes albopictus* hosts, remains completely unknown. Transitioning from the current cellular level to tissue- and organism-level surveillance will be essential to definitively map the ecological prevalence, biogeography, and true natural history of this potential novel genus within global vector ecosystems. More importantly, whether or not our CryoSeekV can be expanded to seek the uncharacterized viruses from the wild raw materials needs further investigation.

Materials and Methods

Cells and Viruses. *Aedes albopictus* C6/36 cells were maintained in Dulbecco's modified Eagle medium (DMEM; Gibco) supplemented with 10% fetal bovine serum (FBS; Gibco) and 1% penicillin-streptomycin at 28°C in a humidified incubator with 5% CO_2. CRFK, BHK-21, Vero and 293T cells were maintained in Dulbecco's modified Eagle medium (DMEM; Gibco) supplemented with 10% fetal bovine serum (FBS; Gibco) and 1% penicillin-streptomycin at 37°C in a humidified incubator with 5% CO_2. AaTLV--IMCAS was discovered in our C6/36 cells by chance. We then purified and amplified the virus for further studies.

Virus Purification. Cell culture supernatants were clarified by centrifugation at 3,000 × g for 10 min at 4°C to remove cellular debris and then passed through a 0.45 μm filter. Clarified supernatants were concentrated by ultracentrifugation at approximately 100,000 × g for 2 to 3 h at 4°C. Pellets were resuspended in PBS (pH 7.4).

For further purification, concentrated samples were layered onto 20% to 60% (wt/vol) sucrose density gradients and centrifuged at 100,000 to 150,000 × g for 2 to 4 h at 4°C. Visible bands were collected fractionally and analyzed by negative-stain electron microscopy. Fractions containing homogeneous 35 nm particles were pooled and used for cryo-EM, RNA extraction, and infectivity assays.

Cryo-EM Sample Preparation Purified virus particles were adjusted to approximately 12 mg/ml. Aliquots (4 µL) were applied to glow-discharged (18 s, 18 mA, 0.39 MPa) Quantifoil Au 300 mesh R2/2 grids. Grids were blotted for 6.5 s with a blot force of −4 at 4°C and 100% humidity using a Vitrobot Mark IV (Thermo Fisher Scientific) and plunge-frozen in liquid ethane.

Cryo-EM data were collected on a Titan Krios transmission electron microscope (TEM) (Thermo Fisher Scientific) operated at 300 kV and equipped with a Falcon 4i direct electron detector. Movies were recorded in counting mode at a nominal magnification corresponding to a calibrated pixel size of 0.97 Å. A total dose of 60 $e^-/Å^2$ was fractionated over multiple frames. In total, 5,160 movies were collected using EPU automated acquisition software (Table. S1)

Image Processing and Reconstruction. Movie frames were motion corrected using Patch Motion Correction (31) in cryoSPARC v4 (17). Contrast transfer function (CTF) parameters were estimated using Patch CTF Estimation (32). Initial particle picking was performed using Blob Picker, followed by particle extraction and two rounds of reference-free 2D classification to remove contaminants and damaged particles.

A total of 97,874 particles were retained after 2D classification. Ab initio three-dimensional reconstruction was performed without symmetry (C1), and the best class containing 54,310 particles was selected for further refinement. Non-uniform refinement (NU-refine) (33) was subsequently performed with icosahedral symmetry (I) imposed. Global and local CTF refinements were iteratively applied, followed by new rounds of non-uniform refinement with I symmetry.

The final reconstruction reached an overall resolution of 2.11 Å based on the gold-standard Fourier shell correlation (FSC) criterion of 0.143 (34). Local resolution estimation and map sharpening were performed in cryoSPARC (17). DeepEMhancer was additionally used for visualization-enhanced post-processing (35).

Model Building and Refinement. For structural identification, density corresponding to a single capsid protomer was segmented in UCSF ChimeraX (36). A polyalanine backbone model was manually traced in Coot (37) and used for structural similarity searches with Foldseek (24).

Automated atomic model building was performed using ModelAngelo (21). Predicted peptide fragments were used for sequence database searches against the National Center for Biotechnology Information (NCBI) nonredundant protein database using the basic local alignment search tool for proteins (BLASTp) (38, 39). Candidate capsid protein sequences were manually inspected and fitted into the density map.

The final atomic model was refined against the cryo-EM map using PHENIX (40) real-space refinement (18) with geometry restraints based on the initial auto-build model. Model quality was assessed using MolProbity (41), Ramachandran statistics

(42), clashscore (41), bond geometry, and map-model FSC.

RNA Extraction and Genome Sequencing. Viral RNA was extracted from purified particles using MagaBio Plus Viral DNA/RNA Purification Kit III according to the manufacturer's instructions. RNA libraries were prepared for NGS, and reads were assembled using de novo strategies.

To fill low-coverage regions, reverse transcription was performed using the HiScript IV 1st Strand cDNA Synthesis Kit with oligo(dT), random primers, or sequence-specific primers. Polymerase chain reaction (PCR) products were amplified using Phanta Max high-fidelity DNA polymerase and subjected to Sanger sequencing. Genome termini were determined by rapid amplification of cDNA ends (RACE).

Quantitative RT-PCR. Viral RNA in culture supernatants was quantified by one-step qRT-PCR. Briefly, supernatants collected from infected cells were clarified by low-speed centrifugation to remove cellular debris and subjected to RNA extraction using the MagaBio Plus Viral DNA/RNA Purification Kit III (Bioer, China) according to the manufacturer's instructions. Purified RNA was eluted in RNase-free water and stored at −80°C until analysis.

Gene-specific primers targeting the viral genome were designed based on the assembled sequence. The finalized TaqMan probe-based assay consisted of the forward primer (5'-CACCTGCATGCGGCTGAC-3'), reverse primer (5'-TCGTCGCCACTGACACAGAC-3'), and a fluorogenic probe (5'-FAM-

CGGCGAGCCCGGCACCTATGACG-TAMRA-3’). qRT-PCR was performed using One Step PrimeScript™ RT-PCR Kit (Perfect Real Time) (Takara, Japan) in 7500 Fast Real-Time PCR System (Thermo Fisher, USA) following the manufacturer’s cycling conditions. Each reaction contained extracted RNA template, forward and reverse primers, enzyme mix, and reaction buffer in a final volume recommended by the supplier. No-template controls and mock-infected controls were included in each run.

Growth Curve Analysis. To determine viral replication kinetics, C6/36, CRFK, BHK-21, Vero and 293T cells were seeded in 12-well plates and cultured to approximately 100% confluence. Cells were inoculated with purified virus and allowed to adsorb for 1 h at room temperature. Following adsorption, inocula were removed, cells were washed once with phosphate-buffered saline (PBS), and fresh maintenance medium was added.

For C6/36 cells, infected cultures were incubated at 28°C, 33°C, or 37°C to assess temperature-dependent replication. At the indicated time points post-infection, aliquots of culture supernatants were collected. Samples were clarified by brief centrifugation and stored at −80°C prior to RNA extraction.

Viral RNA levels in supernatants were quantified by qRT-PCR as described above. Growth curves were generated by plotting viral genome copy numbers versus time post infection. Each condition was performed with at least three biological replicates.

Cytopathic Effect Assay. To evaluate whether infection induced visible cytopathic effects (CPE), C6/36 cells were seeded in multiwell plates and infected with purified virus under the same adsorption conditions described above. Mock-infected cells treated with culture medium alone were included as controls.

Following infection, cells were maintained at the indicated temperature and examined daily using an inverted light microscope. Cellular morphology was documented throughout the observation period, including cell rounding, detachment, syncytium formation, lysis, reduced adherence, and abnormal clustering. Representative images were captured at selected time points.

Phylogenetic Analysis. To determine the evolutionary position of AaTLV within the order *Tymovirales* and evaluate its potential assignment to a novel genus, a robust phylogenetic analysis was performed using full-length viral genomic sequences. Representative sequences encompassing three major established genera within the family *Tymoviridae*—namely *Tymovirus*, *Maculavirus*, and *Marafivirus*—were retrieved from the official International Committee on Taxonomy of Viruses (ICTV) taxonomy database. Four previously documented mosquito-borne *Tymoviridae*-like viruses, including *Culex Tymo-like virus strain Los Angeles* (CTymVL, GenBank accession no. MH188018.1), *Culex Tymoviridae-like virus* (CuTLV, NC_018703.1), *Ek Balam virus* (EkBV, MH822863.1), and *Guachaca virus* (GUAV, OQ286121.1), were incorporated into the dataset. *Potato virus X* (PVX, MT752936.1) was selected as the outgroup to root the tree. A total of 31 full-length nucleotide sequences were compiled.

Multiple sequence alignment (MSA) of the full-length genomes was executed using the MUSCLE algorithm (v5.1) (Edgar 2004) integrated within the Geneious Prime software platform(43). Following global alignment, conserved nucleotide regions encoding the core RNA-dependent RNA polymerase (RdRp) domain were precisely identified based on genomic domain annotations and systematically extracted from the aligned dataset. The extracted RdRp segment alignment was subsequently imported into MEGA X software(44) for refined local alignment optimization and downstream evolutionary modeling.

The best-fitting nucleotide substitution model was determined using the "Find Best DNA/Protein Models" suite in MEGA X. The models were evaluated and ranked according to the Akaike Information Criterion corrected for small sample sizes (AICc)(45). The General Time Reversible model incorporating a Gamma distribution for rate heterogeneity across sites and a fraction of invariable sites (GTR+G+I)(46-48) yielded the lowest AICc value and was selected for phylogenetic reconstruction. A maximum-likelihood (ML) phylogenetic tree(44) was then reconstructed under the GTR+G+I model. Topological reliability and node support were rigorously evaluated using the bootstrap resampling method with 1,000 replicates(49). The finalized phylogenetic tree was visualized and annotated to elucidate the distinct evolutionary trajectory of the AaTLV branch.

Cell culture, Viral infection and RNA Sequencing. *Aedes albopictus* C6/36 cells were inoculated with virus or mock, and total RNA was extracted using TRIzol reagent at 24 and 48 h post-infection and subjected to RNA-seq.

Read Quantification, Filtering and Gene Annotation. Sequencing reads were quantified to a gene-level count matrix against the *Aedes albopictus* reference genome (assembly GCF_001876365.2, canu_80X_arrow2.2) and its NCBI RefSeq gene annotation (doi: 10.1073/pnas.1516410112). Genes with zero counts across all libraries were removed prior to analysis, and low-count genes were additionally excluded during testing by independent filtering on mean normalized expression to maximize the number of genes passing the significance threshold. Because most C6/36 genes carry only LOC identifiers, human-readable gene symbols were assigned where possible by orthology using KEGG Orthology (KO) mapping (PMID: 10592173); a small number of immune-related loci lacking a clear one-to-one ortholog were annotated by their conserved protein domain to the closest human family member (for example, phenoloxidase to the tyrosinase family, TYR; fibrinogen-domain proteins to FGL1/MFAP4).

Normalization and Differential Expression. Differential expression was assessed with a negative-binomial generalized linear model following the DESeq2 framework (PMID: 25516281). Library size factors were estimated by the median-of-ratios method and used to normalize counts. Gene-wise dispersions were estimated by a Cox–Reid-adjusted profile likelihood (PMID: 22287627), a parametric mean–dispersion trend was fitted, and individual estimates were shrunk toward the trend using an empirical-Bayes procedure. For each timepoint, the Virus-versus-Mock coefficient was tested with a

Wald test, and *P* values were adjusted for multiple testing by the Benjamini–Hochberg false-discovery-rate (FDR) procedure with independent filtering. Genes with adjusted $P < 0.05$ and $|\log 2 \text{ fold change}| > 1$ were considered differentially expressed. The pipeline was implemented in Python and validated on simulated negative-binomial data.

Two-factor Time × Infection Interaction Model. To identify genes whose infection response differed between timepoints, counts from all ten libraries were jointly modelled with a two-factor negative-binomial GLM including terms for timepoint, infection and their interaction (~ time + infection + time:infection). Interaction significance was assessed by a likelihood-ratio test comparing the full model to a reduced model lacking the interaction term, with Benjamini–Hochberg FDR correction; genes with a significant interaction and a substantial difference in infection fold change between timepoints were retained.

Functional Enrichment and Gene-set Enrichment Analysis. KEGG pathway and Gene Ontology (Biological Process, Cellular Component, Molecular Function). Over-representation among the DEGs was tested by the hypergeometric test against the background of all expressed, tested genes, with Benjamini–Hochberg correction. In addition, rank-based gene-set enrichment analysis (PMID: 16199517) was performed by ranking genes by their 24 h Wald statistic and evaluating each pathway's member genes for a shift in rank distribution using a Mann–Whitney rank test, yielding a signed enrichment statistic; infection- and immunity-associated KEGG gene sets were tested for directional enrichment in the infected transcriptome.

Functional Protein–Protein Interaction Network. A functional protein–protein interaction network of the 24 h DE proteins was constructed from shared KEGG pathway membership: an edge was drawn between two proteins co-annotated to one or more common pathways, weighted by the number of shared pathways. The network was laid out with the Kamada–Kawai algorithm, node connectivity was summarized as weighted degree, and functional modules were detected by greedy modularity-maximization community detection. Nodes were coloured by log2 fold change and sized by connectivity, and the most connected hubs were labelled.

Dimensionality Reduction and Temporal-Archetype Classification. Sample-level structure was summarized by principal-component analysis (PCA) on the 1,000 most variable genes (log2 normalized counts). A gene-level t-SNE embedding (van der Maaten & Hinton, 2008) of the DEGs was computed from expression across all libraries to visualize temporal classes and fold-change structure. Each DE gene was assigned to a temporal archetype from the sign and magnitude of its infection log2 fold changes at 24 and 48 h and their significance (sustained, transient, late and reversed up/down responders).

Data Availability. The cryo-EM density map and atomic coordinates have been deposited in the EMDB and PDB under accession codes EMD-XXXXX and PDB XXXX, respectively. The complete viral genome sequence is deposited in GenBank under accession number PPXXXXXX. Raw transcriptomic data are available in the NCBI SRA under BioProject PRJNAXXXXXX.

ACKNOWLEDGMENTS. The Cryo-EM related work was completed at the Cryo-Electron Microscopy Facility of the Electron Microscopy Center at Shandong First Medical University. This work was supported by the Prevention and Control of Emerging and Major Infectious Diseases-National Science and Technology Major Project (2025ZD01900500, 2025ZD01901400), the External Cooperation Program of Chinese Academy of Sciences (180GJHZ2023017MI) and the Capital's Funds for Health Improvement and Research (CFH2026-1Q-1201). Xiaoming Li was supported by the Sanming Project of Medicine in Shenzhen (No. SZSM202311023) and the Guangdong High-level Hospital Construction Fund.

References

1. A. A. Zayed *et al.*, Cryptic and abundant marine viruses at the evolutionary origins of Earth's RNA virome. *Science* **376**, 156–162 (2022).
2. A. C. Gregory *et al.*, Marine DNA Viral Macro- and Microdiversity from Pole to Pole. *Cell* **177**, 1109–1123.e1114 (2019).
3. M. Shi *et al.*, Redefining the invertebrate RNA virosphere. *Nature* **540**, 539–543 (2016).
4. S. R. Krishnamurthy, D. Wang, Origins and challenges of viral dark matter. *Virus Res* **239**, 136–142 (2017).
5. M. E. J. Woolhouse, L. Brierley, Epidemiological characteristics of human-infective RNA viruses. *Scientific Data* **5**, 180017 (2018).
6. K. B. Scholthof *et al.*, Top 10 plant viruses in molecular plant pathology. *Mol Plant Pathol* **12**, 938–954 (2011).
7. A. Brun, Vaccines and Vaccination for Veterinary Viral Diseases: A General Overview. *Methods Mol Biol* **1349**, 1–24 (2016).
8. Y. Z. Zhang, M. Shi, E. C. Holmes, Using Metagenomics to Characterize an Expanding Virosphere. *Cell* **172**, 1168–1172 (2018).
9. S. Roux *et al.*, Minimum Information about an Uncultivated Virus Genome (MIUViG). *Nature Biotechnology* **37**, 29–37 (2019).
10. S. Nurk, D. Meleshko, A. Korobeynikov, P. A. Pevzner, metaSPAdes: a new versatile metagenomic assembler. *Genome Res* **27**, 824–834 (2017).
11. N. Fierer, Embracing the unknown: disentangling the complexities of the soil microbiome. *Nature Reviews Microbiology* **15**, 579–590 (2017).
12. W. Kühlbrandt, The Resolution Revolution. *Science* **343**, 1443–1444 (2014).
13. Y. Cheng, Single-particle cryo-EM—How did it get here and where will it go. *Science* **361**, 876–880 (2018).
14. T. Nakane *et al.*, Single-particle cryo-EM at atomic resolution. *Nature* **587**, 152–156 (2020).
15. K. M. Yip, N. Fischer, E. Paknia, A. Chari, H. Stark, Atomic-resolution protein structure determination by cryo-EM. *Nature* **587**, 157–161 (2020).
16. S. H. Scheres, RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**, 519–530 (2012).
17. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods* **14**, 290–296 (2017).
18. P. V. Afonine *et al.*, Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* **74**, 531–544 (2018).
19. C. M. Ho *et al.*, Bottom-up structural proteomics: cryoEM of protein complexes enriched from the cellular milieu. *Nat Methods* **17**, 79–85 (2020).
20. G. Chojnowski *et al.*, findMySequence: a neural-network-based approach for identification of unknown proteins in X-ray crystallography and cryo-EM. *IUCrJ* **9**, 86–97 (2022).
21. K. Jamali *et al.*, Automated model building and protein identification in cryo-EM maps. *Nature* **628**, 450–457 (2024).
22. T. Wang *et al.*, CryoSeek II: Cryo-EM analysis of glycofibrils from freshwater reveals well-structured glycans coating linear tetrapeptide repeats. *Proc Natl Acad Sci U S A* **122**, e2423943122 (2025).

23. T. Wang *et al.*, CryoSeek: A strategy for bioentity discovery using cryoelectron microscopy. *Proc Natl Acad Sci U S A* **121**, e2417046121 (2024).

24. M. van Kempen *et al.*, Fast and accurate protein structure search with Foldseek. *Nature Biotechnology* **42**, 243–246 (2024).

25. M. Sadeghi *et al.*, Virome of > 12 thousand Culex mosquitoes from throughout California. *Virology* **523**, 74–88 (2018).

26. L. Wang *et al.*, Genomic characterization of a novel virus of the family Tymoviridae isolated from mosquitoes. *PLoS One* **7**, e39845 (2012).

27. J. Charles *et al.*, Discovery of a novel Tymoviridae-like virus in mosquitoes from Mexico. *Arch Virol* **164**, 649–652 (2019).

28. K. Laiton-Donato *et al.*, Novel Putative Tymoviridae-like Virus Isolated from Culex Mosquitoes in Colombia. *Viruses* **15** (2023).

29. E. E. Matsumura *et al.*, Citrus sudden death-associated virus as a new expression vector for rapid in planta production of heterologous proteins, chimeric virions, and virus-like particles. *Biotechnol Rep (Amst)* **35**, e00739 (2022).

30. S. Gutiérrez, Y. Michalakis, M. Van Munster, S. Blanc, Plant feeding by insect vectors can affect life cycle, population genetics and evolution of plant viruses. *Functional Ecology* **27**, 610–622 (2013).

31. S. Q. Zheng *et al.*, MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331–332 (2017).

32. A. Rohou, N. Grigorieff, CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216–221 (2015).

33. A. Punjani, H. Zhang, D. J. Fleet, Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat Methods* **17**, 1214–1221 (2020).

34. M. van Heel, M. Schatz, Fourier shell correlation threshold criteria. *Journal of Structural Biology* **151**, 250–262 (2005).

35. R. Sanchez-Garcia *et al.*, DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Commun Biol* **4**, 874 (2021).

36. E. F. Pettersen *et al.*, UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605–1612 (2004).

37. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–2132 (2004).

38. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410 (1990).

39. S. F. Altschul *et al.*, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402 (1997).

40. D. Liebschner *et al.*, Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861–877 (2019).

41. V. B. Chen *et al.*, MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12–21 (2010).

42. G. N. Ramachandran, C. Ramakrishnan, V. Sasisekharan, Stereochemistry of polypeptide chain configurations. *J Mol Biol* **7**, 95–99 (1963).

43. M. Kearse *et al.*, Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).

44. S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* **35**, 1547–1549 (2018).
45. S. Kumar *et al.*, MEGA12: Molecular Evolutionary Genetic Analysis Version 12 for Adaptive and Green Computing. *Mol Biol Evol* **41** (2024).
46. Z. Yang, Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J Mol Evol* **39**, 306–314 (1994).
47. M. Delarue, An asymmetric underlying rule in the assignment of codons: possible clue to a quick early evolution of the genetic code via successive binary choices. *Rna* **13**, 161–169 (2007).
48. S. Tavaré (1986) Some probabilistic and statistical problems in the analysis of DNA sequences.
49. M. J. Sanderson, CONFIDENCE LIMITS ON PHYLOGENIES: THE BOOTSTRAP REVISITED. *Cladistics* **5**, 113–129 (1989).

Figure legends

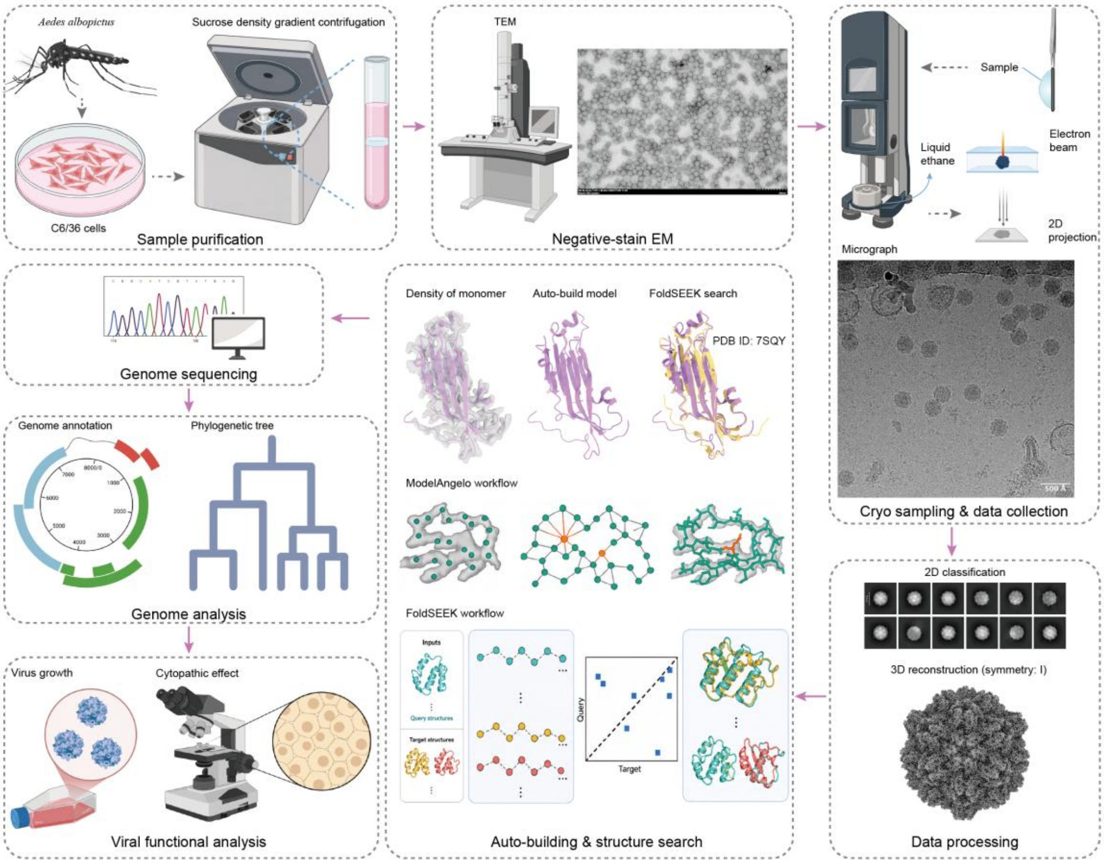


Fig. 1. CryoSeekV workflow for structure-guided discovery and identification of a novel virus. Major steps of the CryoSeekV strategy are illustrated. (1) Virus purification from the culture supernatant of C6/36 cells derived from *Aedes albopictus* by sucrose density gradient centrifugation (SDGC). (2) Negative-stain electron microscopy (EM) analysis of purified particles using an 80-kV transmission electron microscope. (3) Cryo-EM grid preparation and data collection. Scale bar, 500 Å. (4) Cryo-EM image processing, including representative 2D class averages and the final 2.11 Å icosahedrally reconstructed electrostatic potential map. Scale bar, 270 Å. (5) Structural identification using the automated model-building tool ModelAngelo and the structural homology search tool Foldseek. (6) Genome sequencing for validation of the

viral protein sequence and complete viral genome. (7) Genome annotation and phylogenetic analysis. (8) Functional characterization of viral growth and cytopathic effects (CPE). Figure created with BioRender.

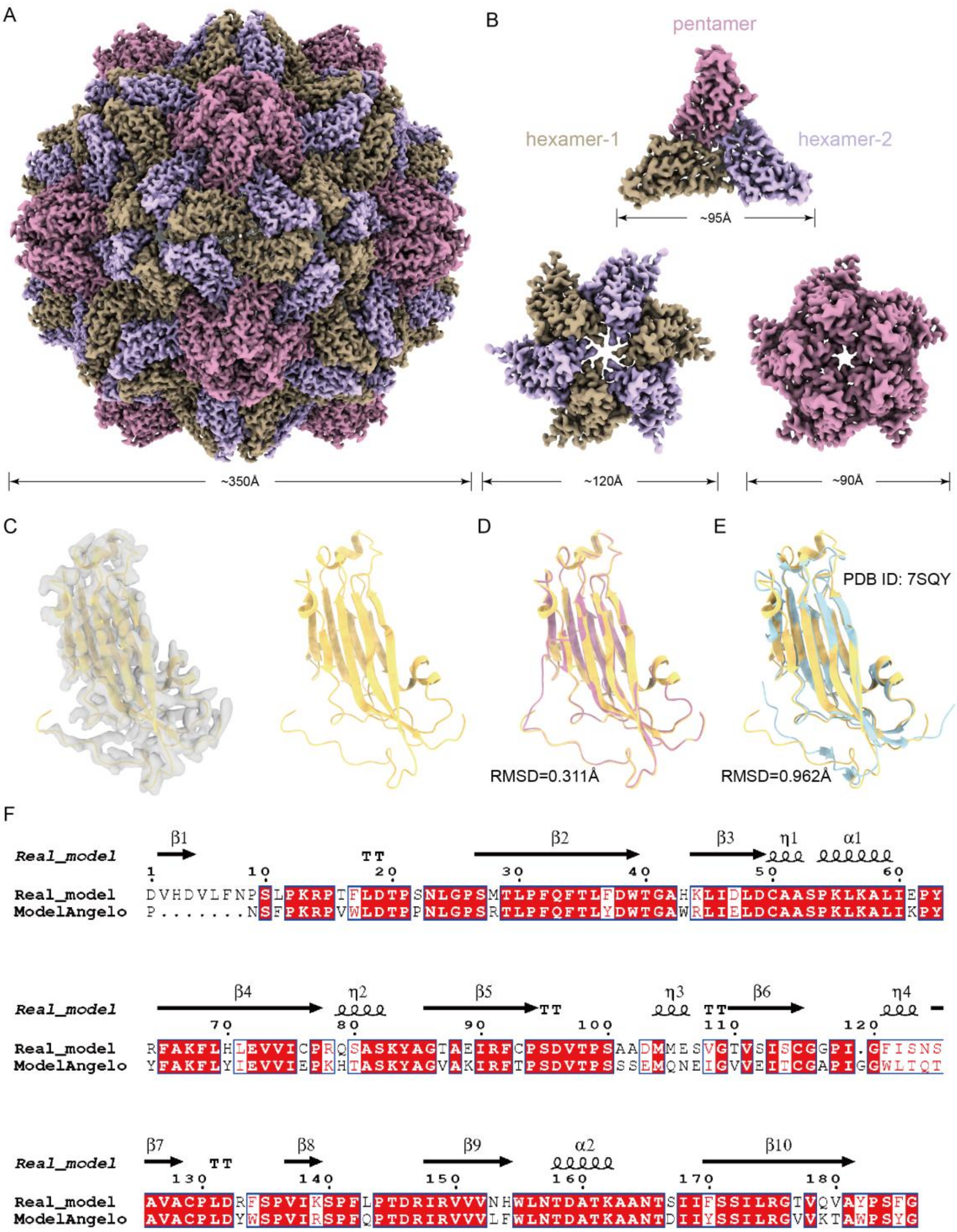
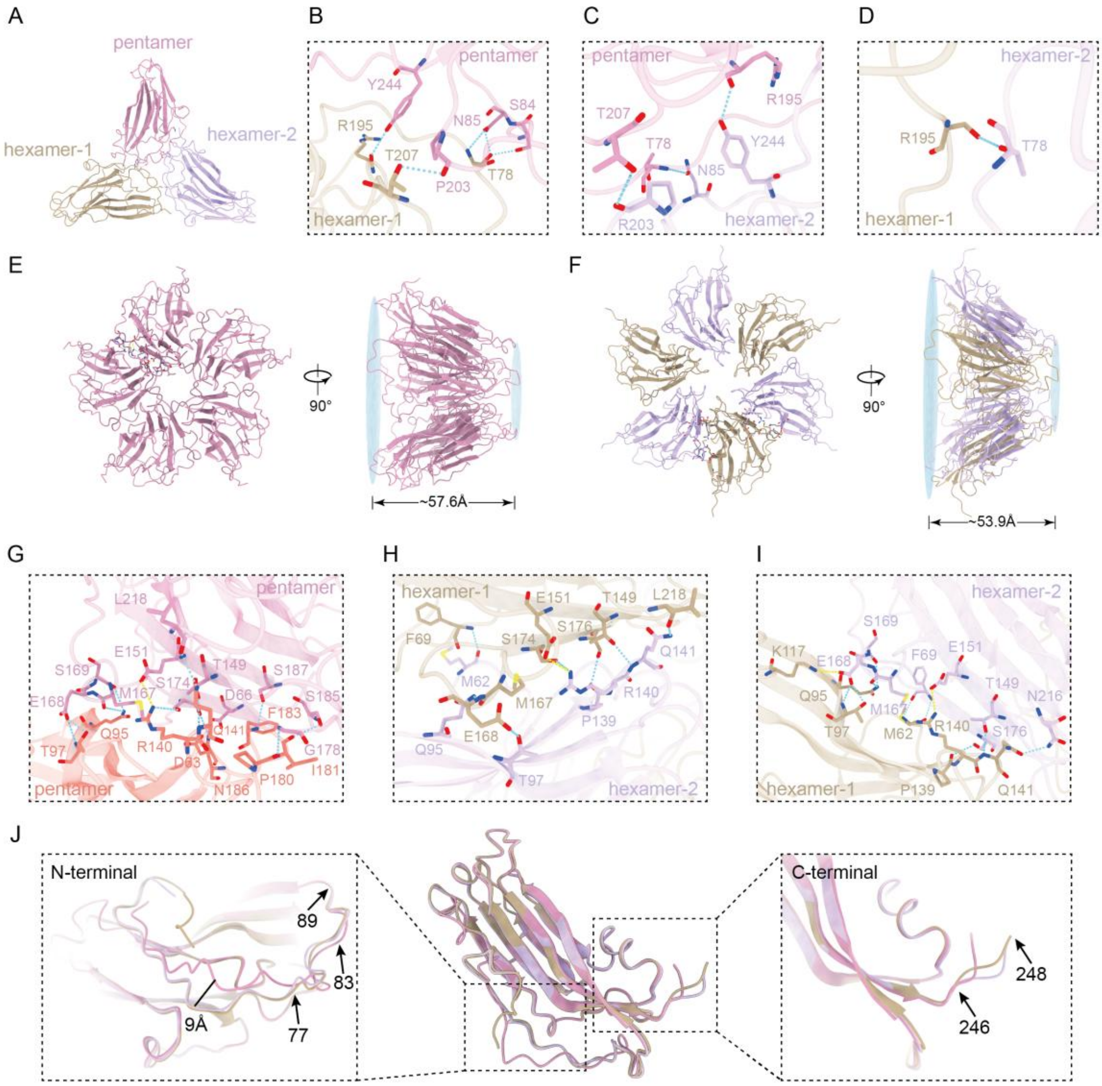


Fig. 2. Structure-based identification of the novel virus. (A and B) The 2.11 Å electrostatic potential map of the virus particle, including the capsid asymmetric unit, pentamer, and hexamer regions, colored according to subunit assignment. Pentameric (pink), hexamer-1 (tan), and hexamer-2 (light purple) subunits constitute the trimeric capsid asymmetric unit. (C) An automated model built into the electrostatic potential

map of a capsid monomer, displayed as a surface-rendered density map. (D) Structural alignment of the automated model (yellow) and the manually refined model (pink), with a root mean square deviation (RMSD) of 0.311 Å. (E) Structural alignment of the automated model (yellow) and the closest structural homolog identified by Foldseek (light blue; PDB ID: 7SQY), with an RMSD of 0.962 Å. (F) Amino acid sequence alignment between the manually refined model and the automated model. The top row indicates the secondary structure of the refined model. Arrows, helices, and "TT" denote β-strands, α-helices, and turns, respectively. Identical residues are highlighted in red, and conserved residues are indicated by blue boxes.



K

Hexamer-1	M62			F69	Q95	T97	K117	P139	R140	Q141	T149	E151	M167
Hexamer-2	M62			F69	Q95	T97		P139	R140	Q141	T149	E151	M167
Pentamer		D63	D66		Q95	T97			R140	Q141	T149	E151	M167

Hexamer-1	E168		S174	S176								L218	
Hexamer-2	E168	S169		S176									N216
Pentamer	E168	S169	S174		G178	P180	I181	F183	S185	N186	S187	L218	

Fig. 3. Structural basis for pentameric and hexameric assembly by an identical capsid protein. (A) Pentameric (pink), hexamer-1 (tan), and hexamer-2 (light purple) subunits forming the trimeric capsid asymmetric unit, shown in cartoon representation. (B to D) Close-up views of the interfaces between pentamer and hexamer-1 subunits, pentamer and hexamer-2 subunits, and hexamer-1 and hexamer-2 subunits, respectively. (E and F) Top and side views of the pentamer (E) and hexamer (F). (G) Close-up view

of the interface between adjacent pentameric subunits during pentamer assembly. (H and I) Close-up views of the interfaces between hexamer-1 and hexamer-2 subunits and between adjacent hexameric subunits during hexamer assembly. Oxygen and nitrogen atoms are colored red and blue, respectively. Blue and yellow dashed lines indicate hydrogen bonds and salt bridges. (J) Structural alignment of pentameric, hexamer-1, and hexamer-2 subunits is shown in cartoon representation, with enlarged views of the N-terminal and C-terminal regions. (K) Residues contributing to pentameric and hexameric assembly in pentameric, hexamer-1, and hexamer-2 subunits.

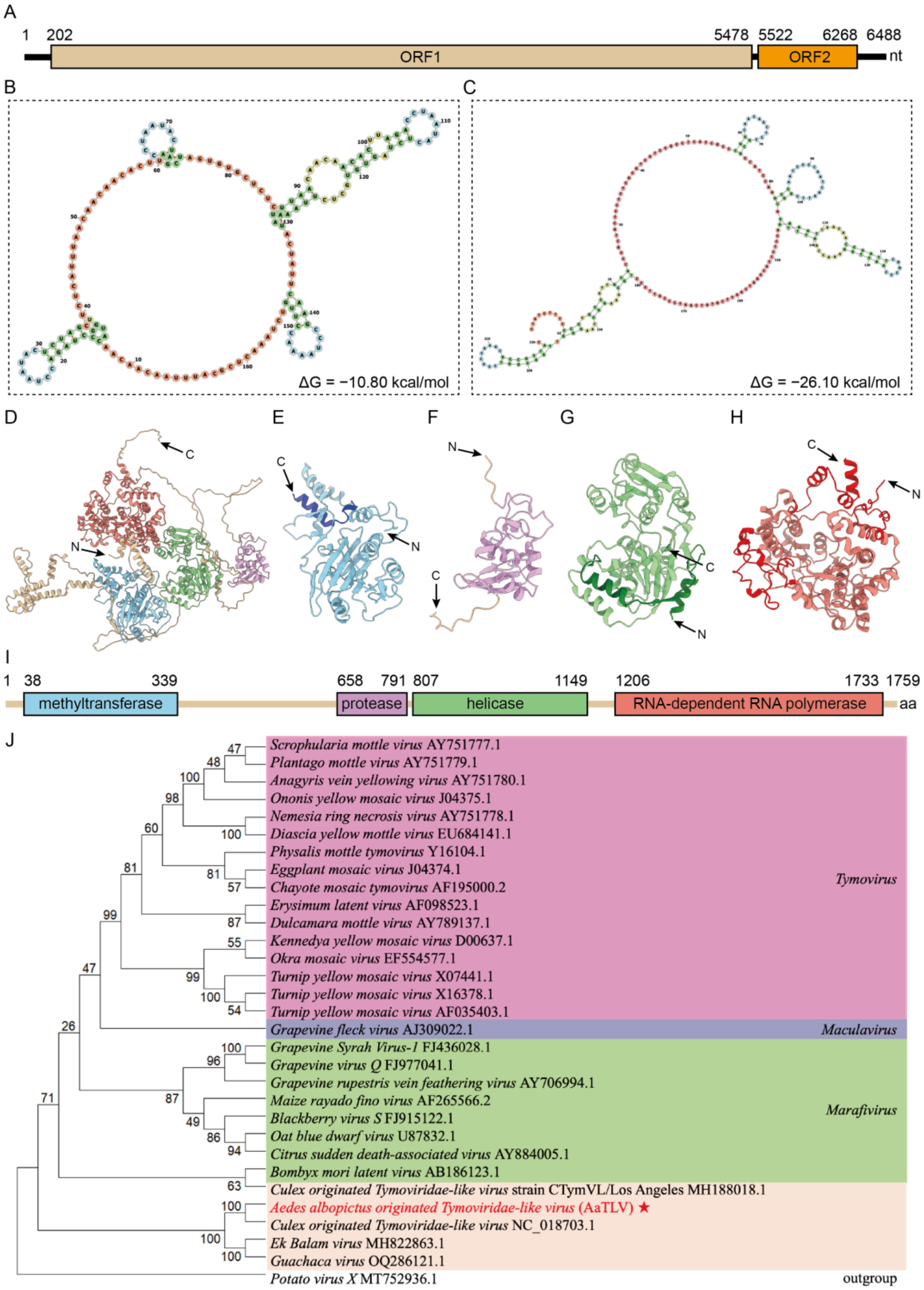


Fig. 4. Genome organization and phylogenetic analysis of *Aedes albopictus*-associated *Tymoviridae-like virus* (AaTLV). (A) Schematic representation of the AaTLV genome. The genome is 6,488 nucleotides in length and contains ORF1 (nt 202

to 5478) and ORF2 (nt 5522 to 6268). (B and C) Predicted stem-loop secondary structures of the 5' untranslated region (UTR) (B; nt 1 to 130) and 3' UTR (C; nt 6269 to 6488). Secondary structures and minimum free-energy values (ΔG) were predicted using RNAfold. (D) Predicted domain organization of the ORF1 polyprotein, including the loop region (tan), methyltransferase (light blue), protease (pink), helicase (green), and RNA-dependent RNA polymerase (RdRp; red). (E) Revised annotation of the methyltransferase domain, with residues 318 to 339 newly assigned (blue). (F) Revised annotation of the protease domain, with residues 653 to 657 and 792 to 804 removed from the predicted domain (tan). (G) Revised annotation of the helicase domain, with residues 807 to 856 and residue 1149 newly assigned (dark green). (H) Revised annotation of the RdRp domain, with residues 1206 to 1314 and 1691 to 1733 newly assigned (dark red). Protein structures shown in panels D to H were predicted using AlphaFold 3. (I) Final annotation of the ORF1 polyprotein. (J) Phylogenetic analysis of viruses in the family *Tymoviridae* based on RdRp amino acid sequences. The phylogenetic tree was reconstructed using the maximum-likelihood method with 1,000 bootstrap replicates. *Tymovirus*, *Maculavirus*, *Marafivirus*, and *Tymoviridae*-like viruses are colored pink, light purple, light green, and tan, respectively.

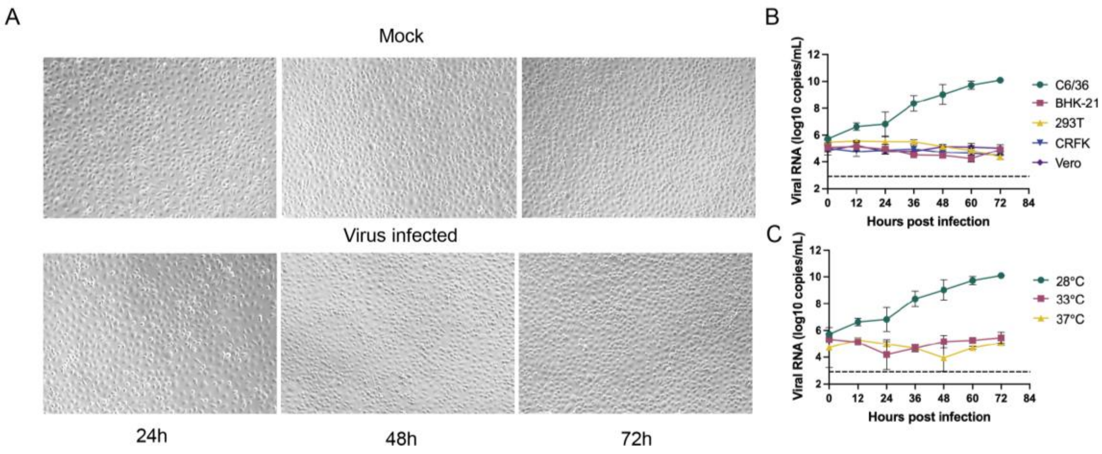


Fig. 5. Functional characterization of AaTLV-IMCAS **infection.** (A) CPE analysis of mock-infected and AaTLV-infected C6/36 cells at 24, 48, and 72 h postinfection. (B) Viral growth kinetics in C6/36, BHK-21, HEK293T, CRFK, and Vero cells. (C) Viral growth kinetics in C6/36 cells cultured at different temperatures. Data are presented as means ± SD from three biological replicates.

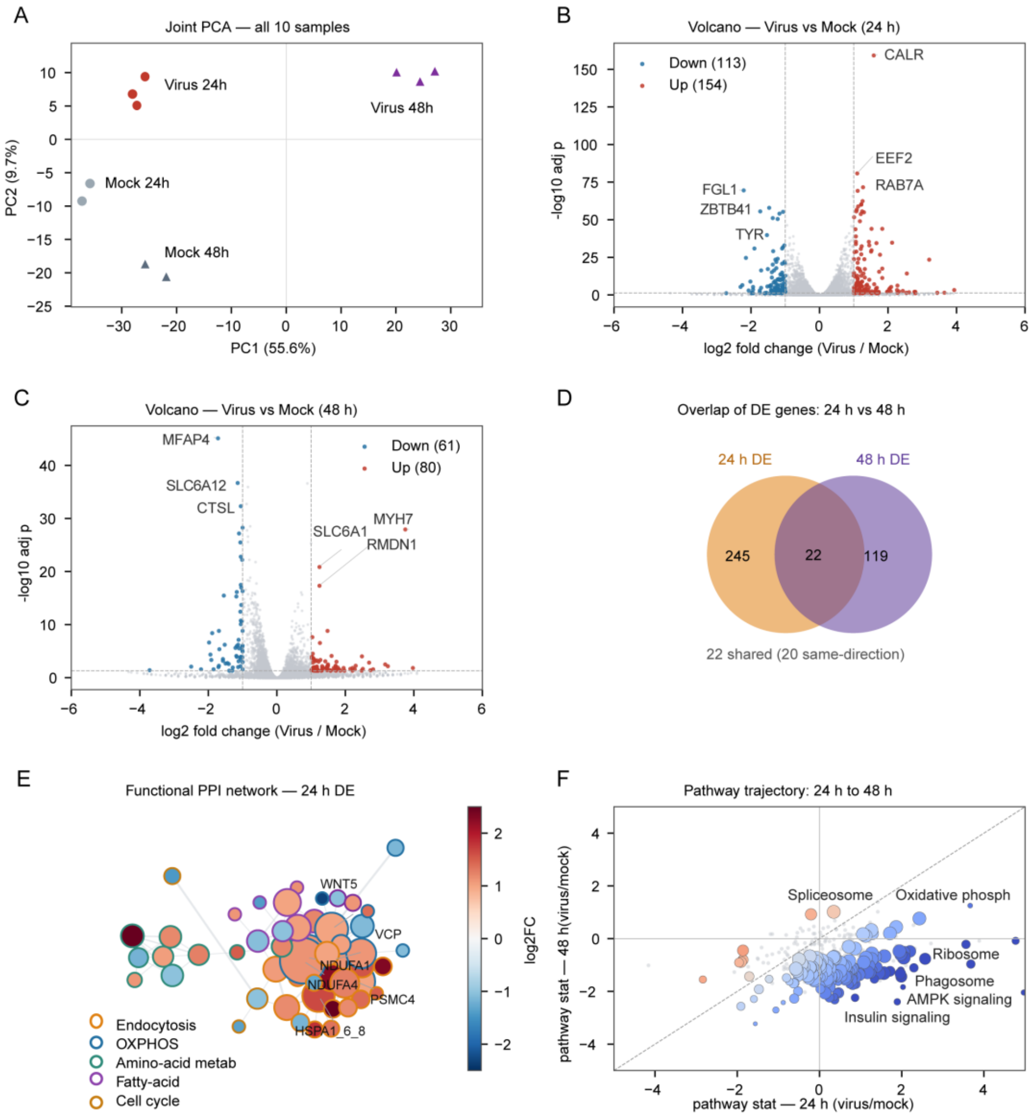


Fig. 6. Temporal transcriptomic response of C6/36 cells to viral infection. (A) Joint principal-component analysis (PCA) of all ten libraries computed on the 1,000 most variable genes; PC1 (55.6% of variance) separates 24 h from 48 h samples and PC2 (9.7%) separates Virus from Mock (circles, 24 h; triangles, 48 h; colour by group). (B) Volcano plot of differential expression for Virus versus Mock at 24 h (267 DEGs: 154 up, red; 113 down, blue); dashed lines mark adjusted P = 0.05 and |log2 fold change| = 1, and selected genes are labelled. (C) Volcano plot for Virus versus Mock at 48 h (141 DEGs; 80 up, 61 down). (D) Venn diagram of DEG overlap between time points: 245

genes unique to 24 h, 119 unique to 48 h, and 22 shared (20 in the same direction). (E) KEGG-pathway-based functional protein–protein interaction network of the 24 h DE proteins; nodes are proteins (fill, log2 fold change; size, weighted degree/connectivity; border colour, functional module), edges denote shared-pathway functional associations, and the most connected hubs are labelled. (F) Pathway-response trajectory: rank-based mean statistic per KEGG pathway at 24 h (x) versus 48 h (y); point size reflects gene-set size and colour the 24 h→48 h change, and the dashed diagonal marks an unchanged response.