

Crystal structure of a tick-borne flavivirus RNA-dependent RNA polymerase suggests a host adaptation hotspot in RNA viruses

Jieyu Yang^{1,2,†}, Xuping Jing^{1,2,†}, Wenfu Yi¹, Xiao-Dan Li¹, Chen Yao^{1,2}, Bo Zhang^{1,3,*}, Zhenhua Zheng^{1,*}, Hanzhong Wang and Peng Gong^{1,3,*}

¹Key Laboratory of Special Pathogens and Biosafety, Wuhan Institute of Virology, Center for Biosafety Mega-Science, Chinese Academy of Sciences, No.44 Xiao Hong Shan, Wuhan, Hubei 430071, China, ²University of Chinese Academy of Sciences, Beijing 100049, China and ³Drug Discovery Center for Infectious Diseases, Nankai University, Tianjin 300350, China

Received October 21, 2020; Revised December 09, 2020; Editorial Decision December 11, 2020; Accepted December 15, 2020

ABSTRACT

The RNA-dependent RNA polymerases (RdRPs) encoded by RNA viruses represent a unique class of nucleic acid polymerases. RdRPs are essential in virus life cycle due to their central role in viral genome replication/transcription processes. However, their contribution in host adaptation has not been well documented. By solving the RdRP crystal structure of the tick-borne encephalitis virus (TBEV), a tick-borne flavivirus, and comparing the structural and sequence features with mosquito-borne flavivirus RdRPs, we found that a region between RdRP catalytic motifs B and C, namely region B-C, clearly bears host-related diversity. Inter-virus substitutions of region B-C sequence were designed in both TBEV and mosquito-borne Japanese encephalitis virus backbones. While region B-C substitutions only had little or moderate effect on RdRP catalytic activities, virus proliferation was not supported by these substitutions in both virus systems. Importantly, a TBEV replicon-derived viral RNA replication was significantly reduced but not abolished by the substitution, suggesting the involvement of region B-C in viral and/or host processes beyond RdRP catalysis. A systematic structural analysis of region B-C in viral RdRPs further emphasizes its high level of structure and length diversity, providing a basis to further refine its relevance in RNA virus-host interactions in a general context.

INTRODUCTION

Flaviviruses are a diverse group of positive-strand RNA viruses with a wide geographic distribution. The majority of flaviviruses are transmitted between hematophagous arthropods and vertebrate hosts, causing human encephalitis or hemorrhagic diseases in severe cases. To date, 53 flavivirus species were listed in the *Flavivirus* genus of the International Committee on Taxonomy of viruses (ICTV) collection (<https://talk.ictvonline.org>). With respect to the transmission mode, ICTV-enlisted flaviviruses can be divided into three categories: mosquito-borne, tick-borne, and vertebrate-specific (1–3). Well-known mosquito-borne flaviviruses include yellow fever (YFV), Japanese encephalitis (JEV), dengue (DENV), West Nile (WNV) and Zika (ZIKV) viruses, which are human pathogens of global concern (4–7). Tick-borne flaviviruses associated with severe human diseases include tick-borne encephalitis (TBEV), Omsk hemorrhagic fever (OHFV), and Kyasanur Forest disease (KFDV) viruses (6,8–10). Vertebrate-specific flaviviruses, also known as No Known Vector (NKV) flaviviruses, are represented by Modoc virus (MODV) isolated from rodents and Rio Bravo virus (RBV) isolated from bats (11). Flaviviruses not yet included in the ICTV collection contain a group of insect-specific viruses mostly associated with mosquitoes, such as cell fusing agent virus (CFAV) and Kamiti River virus (KRV) (12–14).

TBEV belongs to the tick-borne flaviviruses, which are sub-divided into a mammalian group and a seabird group. TBEV is classified as one of the species within the mammalian group and forms the tick-borne encephalitis (TBE) serocomplex together with KFDV, OHFV, Langat virus (LGTV), etc. The TBEV species is divided into three

*To whom correspondence should be addressed. Tel: +86 27 87197578; Fax: +86 27 87197578; Email: gongpeng@wh.iov.cn

Correspondence may also be addressed to Bo Zhang. Email: zhangbo@wh.iov.cn

Correspondence may also be addressed to Zhenhua Zheng. Email: zhengzh@wh.iov.cn

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

sub-types, namely Western European, Siberian and Far Eastern. Similar to other flaviviruses, TBEV possesses a single-stranded, positive-sense RNA genome of ~11 kilobases (kb). The genome contains a single open reading frame (ORF) that is flanked by 5' and 3' untranslated regions (UTRs) (15,16). The ORF encodes a large polyprotein of about 3400 amino acids that is cleaved, through proteolytic processing, into three structural proteins (capsid-C, premembrane/membrane-prM/M, envelope-E) and seven nonstructural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5). NS5 is the largest and most conserved protein encoded by flaviviruses, which is a unique natural fusion of an N-terminal methyltransferase (MTase) and a C-terminal RNA-dependent-RNA polymerase (RdRP) connected by a 10-residue linker (17,18). The NS5 MTase possesses guanylyltransferase (GTase) activity in the formation of the G5'-ppp-5'A linkage, and can catalyze both guanine N7 and 2'-O methylation steps in the capping process (19,20). So far, the crystal structure of the about 260-residue NS5 MTase has been reported in more than ten flaviviruses including the meaban virus from the seabird group of the tick-borne category (19,21). The NS5 RdRP is the central component of the viral genome replication and transcription machinery. However, crystal structures are only available for several mosquito-borne flavivirus NS5 RdRPs (WNV, DENV, JEV, ZIKV and YFV) (17,20). In recent years, a group of invertebrate-associated flavi-like segmented RNA viruses have been identified (22). Among these so-called Jingmenviruses, some tick-associated species have been found capable of causing human infection and diseases (22–25). While the structural proteins of these viruses do not exhibit apparent homology with the flavivirus counterparts, the nonstructural proteins NSP1 and NSP2 show high-levels of similarity with the flavivirus NS5 and NS3, respectively (26). However, the structural information of the Jingmenvirus NSP1 remains unclear.

With an aim to understand the RdRPs of the tick-associated flavi- and related viruses, here we report the crystal structures of the TBEV NS5 MTase and RdRP. While the overall structures are consistent with known flavivirus MTase and RdRP structures in isolation, we found that the region connecting the RdRP catalytic motifs B and C was of particular interest. This inter-motif region B–C exhibits structure and length diversity in not only flaviviruses but also in RNA viruses in general. While manipulation in this region did not have essential impact on the enzymatic activities of flavivirus RdRPs and did not abolish RNA replication, it showed a lethal phenotype in cell-based virus production experiments. We therefore propose that this region may be a host adaptation 'hotspot' in RNA viruses, thus enriching the understandings of the functional diversity of viral RdRPs.

MATERIALS AND METHODS

Cloning and protein production

The wild type (WT, full-length) TBEV NS5 gene within the DNA clone of tick-borne encephalitis virus strain WH2012 (Far Eastern subtype, GenBank accession no. KJ755186.1) was cloned into a pET26b vector. The MTase (residues 7–264) and RdRP (residues 265–891) production plasmids

were made using a site-directed, ligase-independent mutagenesis method (SLIM) with WT pET26b-TBEV-NS5 as the parental plasmid (27,28). Four full-length NS5 mutant plasmids, TBEV SS (short substitution)/LS (long substitution) and JEV SS/LS, with inter-species substitutions, were made using the SLIM method and corresponding full-length constructs as the parental plasmid (17,29). The JEV NS5 variants were produced as an ubiquitin fusion protein and were subsequently processed *in vivo* by a coexpressed Ubp1 protease as described previously (17,30). Plasmids were transformed into *Escherichia coli* strain BL21-CodonPlus(DE3)-RIL or BL21(DE3)-pCG1 for TBEV or JEV NS5 overexpression, respectively (30). Cells growth and protein induction procedures were as previously described except that for TBEV NS5 and its variants the temperature for overnight culture growth was 30°C, the post-induction temperature was 16°C, and induction time was 20 h (17).

Cell lysis, protein purification, and protein storage for full-length, MTase and RdRP constructs were performed as previously described for the JEV NS5 study (17), except that the final protein samples were stored in a buffer with higher NaCl concentration (500 mM) and 10% (v/v) glycerol for TBEV RdRP crystallization (17). The molar extinction coefficients for the NS5 constructs were calculated based on protein sequence using the ExPASy ProtParam program (<http://www.expasy.ch/tools/protparam.html>). The yield was in the range of 3–58 mg (lowest: full-length TBEV NS5; highest: TBEV MTase) of pure protein per liter of bacterial culture.

Protein crystallization and crystal harvesting

Lamellar or prismatic crystals of TBEV MTase were obtained within 2 weeks by sitting drop vapor diffusion at 16°C using 8–10 mg/ml protein. Normally, 0.5 µl protein was mixed with 0.5 µl reservoir solution. Best MTase crystals were obtained in 0.085 M HEPES (pH 7.5), 8.5% (w/v) PEG8000, 15% (v/v) glycerol. TBEV RdRP crystal screening was performed at 16°C and 5°C by sitting-drop vapor diffusion, and 0.5 µl of protein solution was mixed with 0.5 µl of reservoir solution. Micro-crystals appeared under several HEPES-containing conditions. The conditions were optimized, and diffraction-quality crystals were obtained in 0.1 M HEPES (pH 7.0), 0.2 M sodium thiocyanate, 15% (v/v) pentaerythritol propoxylate (5/4 PO/OH) and 5 mM MgSO₄. Best diffracting crystals were obtained by crystal dehydration. MTase and RdRP crystals were flashly cooled and stored in liquid nitrogen prior to data collection.

Structure determination and analyses

Single crystal X-ray diffraction data were collected at the Shanghai Synchrotron Radiation Facility (SSRF) beamline BL17U1 (wavelength = 0.9792 Å, temperature = 100 K). At least 180° of data were typically collected in 0.2 oscillation steps. The reflections of the MTase and RdRP data were integrated, merged and scaled using XDS and HKL2000, respectively (Table 1) (31,32). The initial structure solutions of the MTase and RdRP were obtained using the molecular replacement program PHASER using coordinates derived

Table 1. X-ray diffraction data collection and structure refinement statistics

PDB code	MTase – 7D6M	RdRP – 7D6N
Data collection^a		
Space group	<i>P1</i>	<i>P2₁2₁2₁</i>
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	35.0, 49.5, 85.3	176.0, 235.2, 327.4
α , β , γ (°)	79.7, 82.8, 77.1	90, 90, 90
Resolution (Å) ^b	50.0–1.89 (1.94–1.89)	50.0–3.17 (3.27–3.17)
<i>R</i> _{merge}	0.055 (0.19)	0.082 (0.64)
<i>R</i> _{meas}	0.077 (0.27)	0.098 (0.77)
CC _{1/2}	0.991 (0.639)	0.993 (0.628)
I / σ I	8.8 (3.0)	13.9 (1.9)
Completeness (%)	88.1 (84.7)	98.3 (99.5)
Redundancy	1.9 (1.9)	3.2 (3.3)
Refinement		
Resolution (Å)	50.0–1.90	50.0–3.17
No. reflections	38,258	225,329
<i>R</i> _{work} / <i>R</i> _{free} ^c (%)	22.3/27.0	19.5/23.3
No. atoms		
Protein	3,998	45,316
Ligand/ion/water	54/–/168	–/30/–
B-factors (Å ²)		
Protein	29.0	65.8
Ligand/ion/water	25.2/–/30.6	–/69.4/–
R.m.s. deviations		
Bond lengths (Å)	0.007	0.011
Bond angles (°)	0.922	1.326
Ramachandran stat. ^d	92.1/7.9/0.0/0.0	86.1/13.4/0.3/0.2

^aOne crystal was used for data collection for each structure.

^bValues in parentheses are for highest-resolution shell.

^c5% of data are taken for the *R*_{free} set.

^dValues are in percentage and are for most favored, additionally allowed, generously allowed and disallowed regions in Ramachandran plots, respectively.

from Meaban virus (MEAV) MTase (PDB entry 2OXT) and JEV NS5 (PDB entry 4K6M) structures as search models, respectively. Manual model building and structure refinement were done using Coot and PHENIX, respectively (33,34). The 3500 K composite simulated-annealing omit $2F_o - F_c$ electron density maps were generated using CNS (35). Due to the imperfection of diffraction of the MTase crystal, the resolution of data was conservatively cut at 1.9 Å with an *R*_{merge} value of 0.19 for the highest resolution shell (Table 1). For the RdRP structure refinement, non-crystallographic symmetry was applied in early rounds but released in later rounds.

Protein superimpositions were done using the maximum likelihood based structure superpositioning program THESEUS (36). For RdRP region B-C focused superimpositions, TBEV NS5 residues 600–669, not the entire RdRP module, were used as the reference region. The pdb entries used in the MTase superimposition are: 2OXT (MEAV), 2OY0 (WNV), 2PX2 (MVEV), 2WA1 (MODV), 3ELU (WESSV), 3EVA (YFV), 3GCZ (YOKV), 4CTJ (DENV3), 4K6M (JEV) and 5M5B (ZIKV). The pdb entries used in RdRP superimposition are: 2HFZ (WNV), 2J7U (DENV3), 5U0C (ZIKV), 5Y6R (CSFV NS5B), 3MWV (HCV NS5B), 1RA6 (PV 3D^{pol}), 1SH0 (NV RdRP), 3AVX (Q β RdRP-(EF-Tu)-(EF-Ts) complex), 6NUS (SARS-CoV nsp12-nsp8 complex), 5A22 (VSV L), 6QPF (IAV PA-PB1-PB2 complex), 6U5O (HMPV L-P complex), 5AMQ (LACV L-RNA complex), 1HHS (bac-

teriophage ϕ 6 RdRP), 5I61 (hPBV RdRP) and 6TY8 (Bm-CPV RdRP).

RNA preparation

Enzyme-based preparation of the 33-nucleotide (nt) RNA template (T33) was as previously described (37,38). The 30-nt template (T30) and 10-nt RNA primer (P10) were chemically synthesized (Integrated DNA Technologies) and the T30 was purified as previously described (39). For primer-dependent polymerase assays, the annealing procedures of the T33/P10 construct were as previously described (38). For dinucleotide-driven polymerase assays, T30 was annealed with a GG dinucleotide primer bearing a 5'-phosphate (P2, Jena BioSciences) at a 1:5 molar ratio following previously described annealing procedures (39).

In vitro polymerase assays

For the primer-dependent polymerase assay, a 20- μ l reaction mixture containing 6 μ M NS5 or its variant, 4 μ M T33/P10 construct, 50 mM MES (pH 6.5), 20 mM NaCl, 5 mM MnCl₂, 2 mM MgCl₂, 5 mM tris (2-carboxyethyl) phosphine (TCEP), 300 μ M ATP, and 300 μ M UTP, 0.01% (v/v) nonidet P 40 (NP40) was incubated at 40°C for 1, 2, 5 min before being quenched with an equal volume of stop solution (95% (v/v) formamide, 20 mM EDTA (pH 8.0), and 0.02% (w/v) xylene cyanol).

For the P2-driven polymerase assay, a 20- μ l reaction mixture containing 6 μ M NS5 or its variant, 4 μ M T30/P2 construct, 50 mM Tris-HCl (pH 7.5), 20 mM NaCl, 5 mM MnCl₂, 2 mM MgCl₂, 5 mM TCEP, 300 μ M ATP and 300 μ M UTP, 0.01% (v/v) *n*-dodecyl- β -D-maltoside (DDM) was incubated at 30°C for 15, 30, 45 or 60 min. The reaction was quenched by an equal volume of stop solution and the mixture was heated at 100°C for 45–60 s. The procedures for denaturing polyacrylamide gel electrophoresis (PAGE), gel staining, and quantitation were as previously described (38).

Cells and antibodies

Baby hamster kidney cells (BHK-21) (American Type Culture Collection (ATCC), CCL-10) were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 100 units/ml of penicillin and 100 μ g/ml of streptomycin. C6/36 cells were grown in RPMI 1640 medium supplemented with 10% FBS, 100 units/ml of penicillin, and 100 μ g/ml of streptomycin. The BHK-21 cells and C6/36 cells were maintained in 5% (v/v) CO₂ at 37°C and 28°C, respectively. By immunizing rabbits with purified TBEV NS5 protein, rabbit-derived polyclonal antibodies against TBEV NS5 protein in the antiserum were collected and purified from immunized rabbit to detect the production of TBEV protein. Monoclonal antibody (4G2) against JEV envelope protein (E) was used to detect viral protein production of JEV. Fluorescein isothiocyanate (FITC)-conjugated goat anti-mouse IgG was used as secondary antibody.

Construction of mutant genome-length cDNA and/or replicon cDNA of TBEV and JEV

Using the full-length infectious clones of the TBEV WH2012 strain and pACYC-JEV-SA14 and the related TBEV replicon as the backbones (40), the TBEV_SS, JEV_LS, JEV_SS mutant clones were constructed. The NS5 fragments bearing the substitution were generated by overlapping PCR and were introduced to the full-length TBEV and JEV clones through XbaI/XhoI and XbaI/SalI restriction sites, respectively. All constructs were verified by DNA sequencing.

Immunofluorescence assays

Plasmids carrying TBEV WT and SS genomic and related sequences were transfected into BHK-21 cells using the Lipofectamine 3000 Reagent according to the manufacturer's protocols. Three days post transfection, the supernatant was collected and then used to infect BHK-21 cells. Cells were collected at 1, 2, 3 days post infection and were fixed with cold 4% paraformaldehyde at room temperature (r.t.). Virus replication was detected using anti-NS5 rabbit-derived polyclonal antibodies and FITC-conjugated goat anti-rabbit IgG as primary and secondary antibodies, respectively. The nuclei were stained with hoechst33258 (Beyotime Biotechnology) staining solution.

The infectious clones of JEV (WT, LS and SS) were linearized with XhoI and then subjected to *in vitro* transcription using a T7 *in vitro* transcription kit (Thermo Fisher Scientific). Approximately 1 µg of transcribed recombinant genomic RNAs were transfected into BHK-21 cells (mammalian) and C6/36 (mosquito) with reagent DMR1E-C (Invitrogen). Then the cell slides were fixed in cold (−20°C) 5% (vol. to vol.) acetone in methanol at r.t. for 10 min. After washing three times with PBS (pH 7.4), the fixed cells were subjected to an immunofluorescence assay (IFA) with 4G2 monoclonal antibody to examine JEV envelope (E) production.

Luciferase assay

According to the manufacturer's protocol, the TBEV replicon and related mutant plasmids were transfected into three 24-well plates using the Lipofectamin 3000 Reagent. Samples were collected at 1, 2 and 3 days time points. The transfected BHK-21 cells were lysed with 100 µl 1× Lysis Buffer (Renilla-Glo Luciferase Assay System from Promega), and stored at −80°C for subsequent luciferase activity analysis. After collecting all samples, the Renilla luciferase (RLuc) activity was measured by mixing 10 µl of cell lysate with 50 µl of substrate solution. All luciferase assays were performed in triplicates.

RT-qPCR analysis

According to the manufacturer's protocol, RNA was extracted from BHK-21 cells and their supernatants using the TRIzol Reagent (Invitrogen), and then the Tiangen's FastKing gDNA Dispelling RT SuperMix reverse transcription kit was used to reverse transcribe 1 µg of total RNA into cDNA. The cDNA was amplified following a

SYBR Green-based quantitative PCR protocol (Bio-Rad) using a real-time quantitative PCR system (Bio-Rad CFX) that requires an activation at 95°C for 3 min, followed by 40 amplification cycles, with 10 s at 95°C followed by 45 s at 60°C for each cycle. Amplification in a universal pair of primers (TBEV qPCR-F: gagagaaggaaaacaggtgc, and TBEV qPCR-R: cccatcactcctgtgtcac) were used to amplify the region spanning nucleotides 2898–3013 of TBEV (corresponding to a segment in the NS1 gene), which is conserved in all TBEV strains. The RNA copies were calculated based on the standard curve composed of 10-fold serial dilutions of the standard infectious clone plasmid DNA.

Statistical analysis

Data of the luciferase assay and RT-qPCR are presented as means ± standard deviations (SD). Student's t-test was used to assess statistical significance, with significance defined by *P* value <0.05 (*), <0.01 (**) or <0.001 (***)

RESULTS

The overall structure of TBEV NS5 MTase and RdRP are largely consistent with known flavivirus NS5 structures

With an aim to understand the structural features of NS5 RdRP from tick-borne flaviviruses and related species, we attempted to crystallize the full-length TBEV NS5 but did not obtain any hits in the screening. Truncated forms of TBEV NS5 were also tested in crystallization trials. Two of these forms, containing the MTase and RdRP modules, respectively, were crystallized, yielding diffraction datasets for structure determination. The crystal structure of TBEV NS5 MTase was solved at 1.9 Å resolution in space group P1 by molecular replacement, with two MTase molecules in each crystallographic asymmetric unit adopting highly consistent conformation (root-mean-square deviation (RMSD) value for all superimposable α-carbon atoms of 0.3 Å) (Table 1 and Supplementary Figure S1). Comparing with previously reported flavivirus MTase structures in isolation, the TBEV MTase adopts the same Rossmann fold with identical topology and very similar conformation, with corresponding RMSD values of 0.6–1.2 Å (chain A of the TBEV structure as the reference; 96–99% residue coverage) (Figure 1A). An *S*-adenosyl-L-methionine (SAM) molecule, the methyl donor of the methylation reaction, was bound in its binding pocket of each MTase.

The structure of TBEV NS5 RdRP was solved at 3.2 Å resolution in space group *P*₂₁₂₁ by molecular replacement, with ten RdRP molecules in each crystallographic asymmetric unit assembled in a dimer of pentamer fashion (Table 1 and Supplementary Figure S2, A–D), even though both the full-length NS5 and RdRP are monomeric in solution (Supplementary Figure S2E). Moreover, there is no conserved inter-molecular interface documented in available crystal forms of full-length and RdRP-only NS5 crystal structures. The packing among these dimer of pentamers resulted in spacious solvent channels in the crystal lattice, likely accounting for the poor-diffracting feature of the crystal (Supplementary Figure S2, A–D). The RMSD values are 0.4–0.5 Å (98–100% residue coverage)

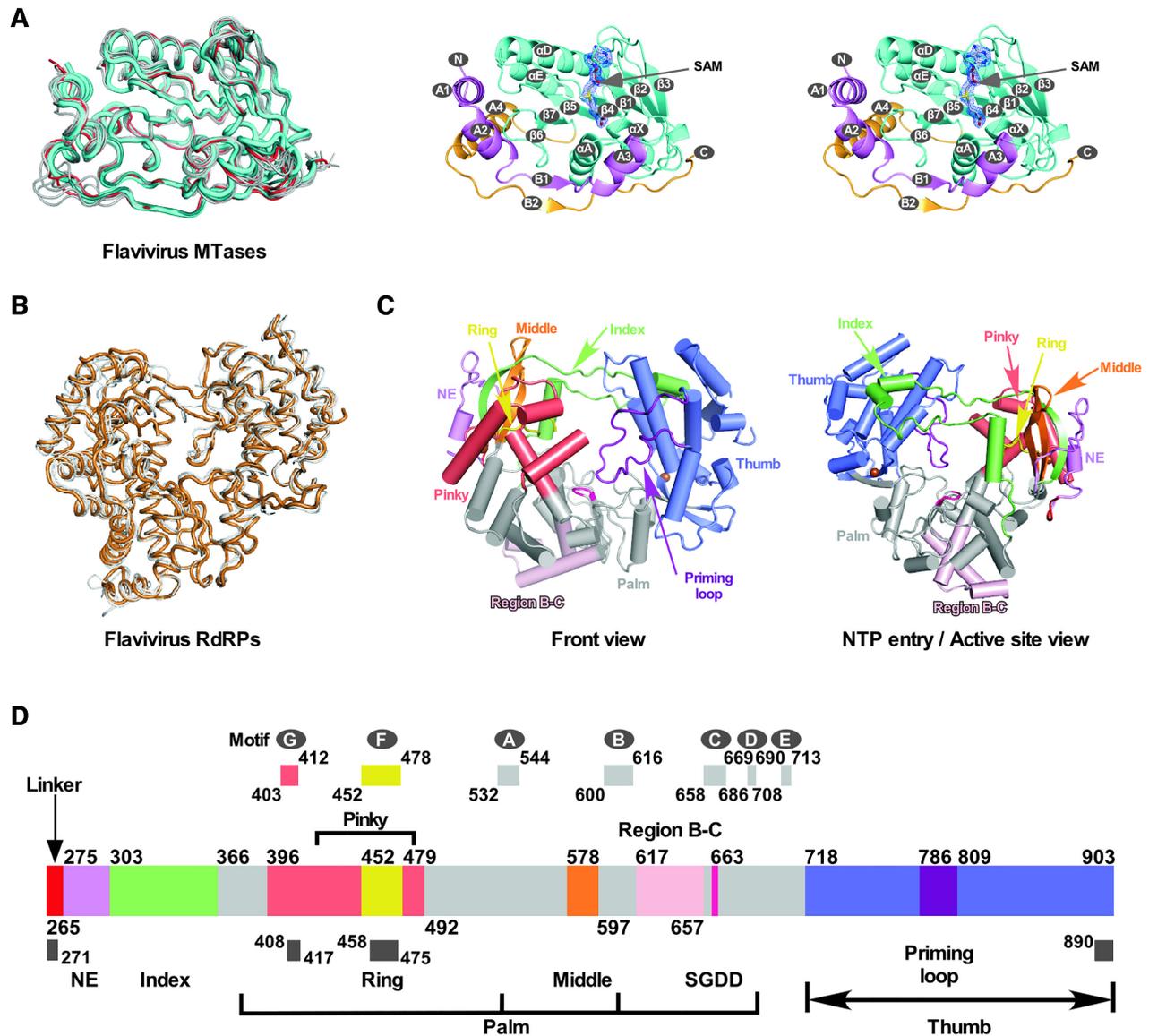


Figure 1. Global views of TBEV NS5 MTase and RdRP structures. (A) Superimposition of the MTases of flaviviruses (left) and stereo-pair view of TBEV MTase in complex with SAM (right). Coloring scheme: left – TBEV MTase in cyan, MEAV MTase in dark turquoise, the MODV and YOKV MTase in red, and mosquito-flavivirus MTases in gray; right – subdomain I in violet, subdomain II in cyan, subdomain III in bright orange, SAM in green and zinc ions (spheres) in brown. The 3500 K composite SA-omit $2F_o - F_c$ electron density map of SAM was overlaid (contoured at 1.0σ). (B) Superimposition of flavivirus RdRPs as cartoon representation. Coloring scheme: TBEV RdRP in dark orange, and mosquito-flavivirus (DENV, WNV) RdRPs in gray. (C) Global views of TBEV RdRP crystal structure. Structure of TBEV RdRP construct shown in orientations viewing into the front channel (left) and NTP entry channel (right), respectively. Coloring scheme: linker in red, RdRP palm in gray, N-terminal extension (NE) in violet, index finger in green, middle finger in orange, ring finger in yellow, pinky finger in light red, thumb in slate, priming element in purple, region B–C in light pink, motif C signature sequence SGDD in magenta. The numbers defining the residue ranges of each element are shown. (D) A color-coded bar of functional or structural elements of TBEV RdRP. The coloring scheme is the same as in panel C, and the unresolved regions in pinky/ring and at both termini are indicated by dark gray bars.

among these chains and $0.9\text{--}1.4 \text{ \AA}$ (74–93% residue coverage) among the RdRP-only flavivirus NS5 structures for all superimposable α -carbon atoms (both using chain A of the TBEV structure as the reference) (Figure 1B and Supplementary Figure S2F). The TBEV structure adopts the encircled human right hand architecture that is RdRP-unique, with palm, fingers, and thumb domains surrounding the polymerase active site (Figure 1B–D). Consistent with available flavivirus RdRP-only structures, part of the ring and

pinky finger subdomains (residues 458–475 in ring and 408–416 in pinky for the TBEV structure) is disordered (20,41). The dynamics in the ring/pinky fingers, and in some cases also in index finger are also observed in full-length NS5 structures (18,39). To date, two types of full-length NS5 global conformations have been identified by crystallography. The conformation represented by the JEV NS5 structure typically has a completely folded fingers domain stabilized by intra-molecular MTase-RdRP interactions fea-

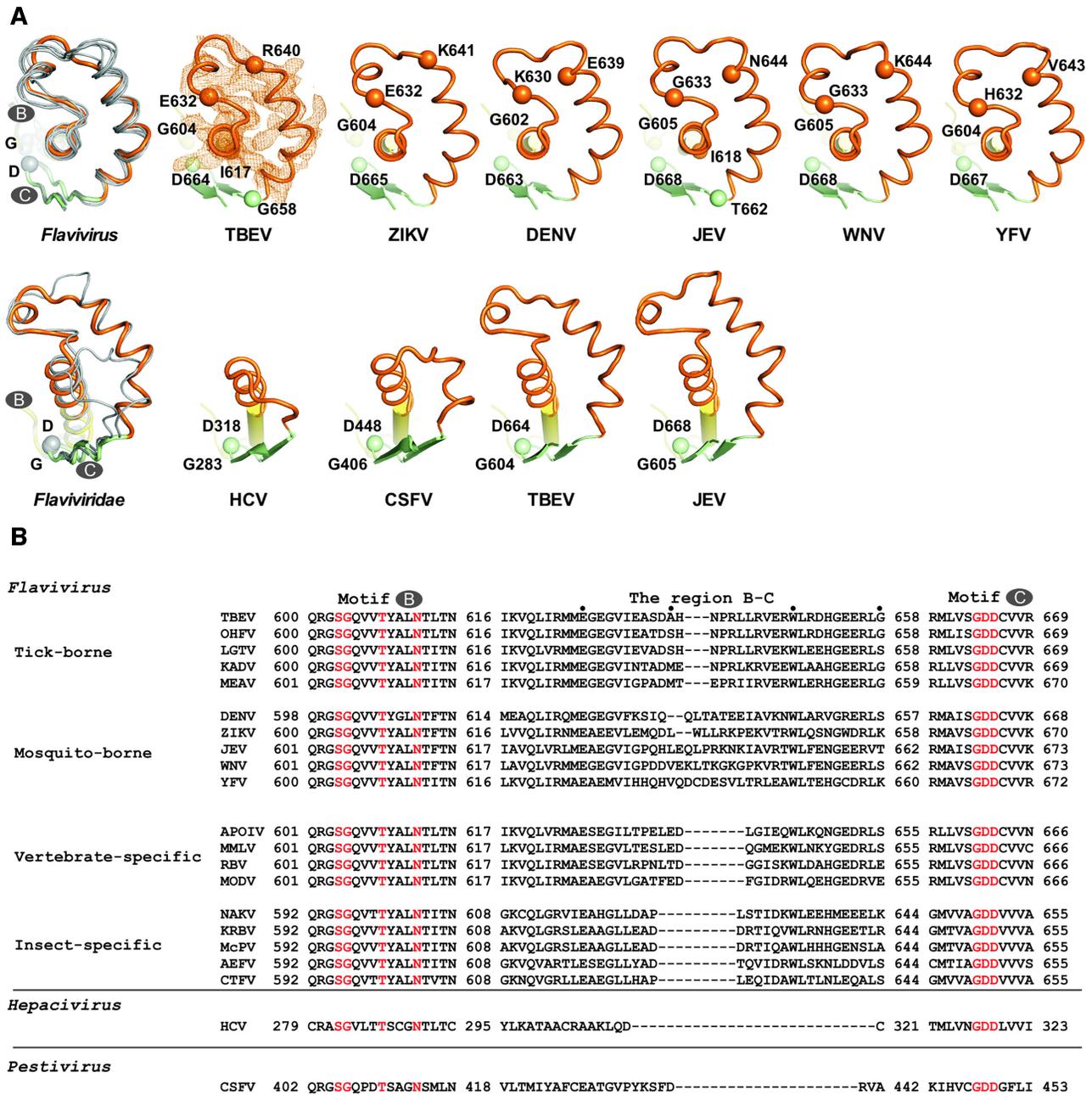


Figure 2. The structure and sequence diversity of region B-C of the *Flaviviridae* NS5 RdRPs. (A) Structural comparison of the motifs B-C region of *Flaviviridae* NS5s. In the first row, TBEV is from tick-borne flaviviruses, the other four are from mosquito-borne flaviviruses. In the second row, TBEV and JEV are from *Flavivirus* genus, and HCV and CSFV represent other two genera of the *Flaviviridae* family. Coloring scheme: motif B in yellow, the region B-C in orange, motif C in light green. Two absolutely conserved residues (the G in the SG sequence of motif B and the first D in XGDD of motif C), and the two end residues of the structurally most diverse part of region B-C are shown as spheres. The α -helix in motif B is shown as a cylinder. The 3500 K composite SA-omit 2Fo - Fc electron density map of TBEV region B-C was overlaid (contoured at 1.0 σ). (B) Structure-assisting sequence alignment of the motifs B-C region of *Flavivirus* genus NS5s. Conserved active site residues are highlighted in red. Depending on the host ranges, viruses in the *Flavivirus* genus are divided into the four groups: the mosquito-borne group, the tick-borne group, the insect-specific group, and the vertebrate-specific group.

tured by a set of hydrophobic interactions (17), and the tip of the ring finger is involved in the interface interactions. By contrast, the conformation represented by the DENV serotype 3 (DENV3) NS5 structure has the RdRP fingers domain largely consistent with the RdRP-only structures (18). The MTase forms a different and primarily polar set of intra-molecular interactions with RdRP in the DENV3

conformation, which does not involve the ring finger tip. The conformational similarity between the TBEV RdRP and the mosquito-borne flavivirus RdRPs nevertheless supports the conservation of NS5 structure and conformation. However, whether the tick-borne flavivirus NS5 can indeed adopt the JEV- and DENV3-like global conformations requires further structural evidence.

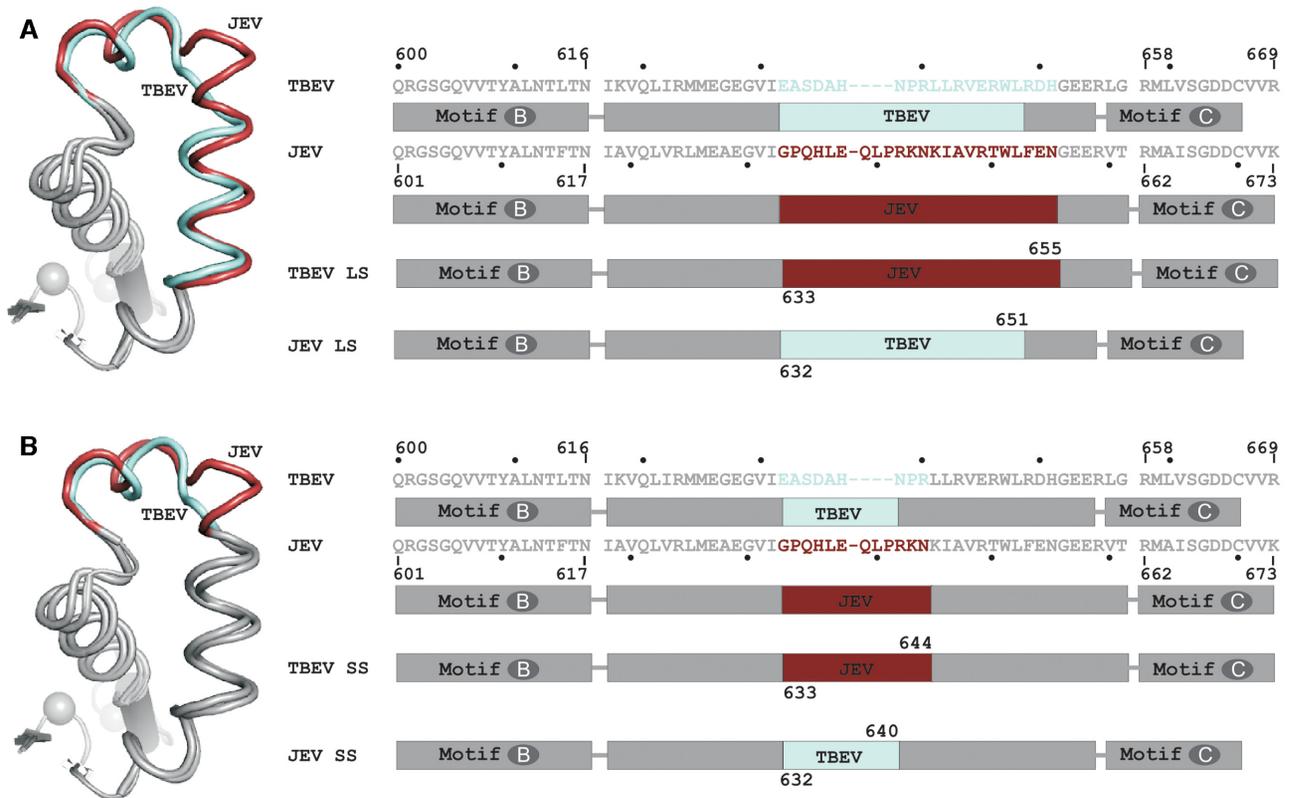


Figure 3. The design of the inter-virus region B-C substitutions in TBEV and JEV backbones. (A) The design of region B-C mutants with substitution of the sequence-variable segment (LS: long substitution). Left: The superimposition of TBEV and JEV motifs B-C region. Right: The motifs B-C region sequences of WT and LS constructs. (B) Left: The design of region B-C variants with substitution of the structurally most divergent segment (SS: short substitution). Right: The motifs B-C region sequences of WT and SS constructs. The TBEV and JEV segments used in substitution were highlighted in cyan and red, respectively.

The region between RdRP motifs B and C exhibits structure and sequence diversity and may have host adaptation-related features

While examining the structural difference between TBEV RdRP and other flavivirus RdRPs, the region between RdRP catalytic motifs B and C (hereinafter termed ‘region B-C’) drew our attention. Motifs B and C are among the most conserved regions of viral RdRPs. Forming a β -hairpin substructure, motif C contains an absolutely conserved aspartic acid residue (the first ‘D’ in the RdRP signature sequence XXDD) in the hairpin region that participates in catalytic divalent metal ion coordination. Motif B is consisted of an N-terminal loop region and a C-terminal helical region, and contains an absolutely conserved glycine residue in the loop (the ‘G’ in a highly conserved SG sequence) that plays an important role in NTP recognition (42,43). Region B-C starts in the central part of the long helix that harbors the motif B helical region and ends at the beginning of the N-terminal β -strand in motif C (Figure 2 and Supplementary Figure S3). This region resides at the bottom of the RdRP palm and is solvent accessible, and therefore has the potential to interact with viral and/or host factors or to mediate processes other than RNA synthesis. Compared to its neighboring motifs, region B-C exhibits relatively high structural variance and dynamics (Supplementary Figure S3, B-C). We previously

carried out a structure and function survey of RdRP from positive-strand RNA viruses and found that the length of region B-C varies to a great extent, with a length difference greater than 60 residues between the longest and the shortest (44). In flaviviruses, the length of this region is about 40 residues and exhibits length and sequence diversity between different groups. The TBEV RdRP region B-C is about 2–3 residues shorter than its mosquito-borne flavivirus counterparts (Figure 2B). If compared with RdRP structures of known mosquito-borne flaviviruses, residues 632–640 of TBEV RdRP or 633–644 of JEV RdRP represent the structurally most diverse part of this region (Figure 2A, top panel). Comparing to HCV and CSFV NS5B proteins, two representative *Flaviviridae* RdRPs, region B-C of flavivirus RdRP is about 15–25 residues longer (Figure 2A, bottom panel).

Inter-species region B-C substitutions do not abrogate RdRP catalytic activities

In order to further explore the role of region B-C in flaviviruses, we designed a set of four inter-species substitution mutations in TBEV and JEV NS5s. The residue ranges chosen for substitution corresponding to the solvent exposed part of region B-C (TBEV NS5 632–651) or the structurally most different stretch (632–640), resulting in two pairs of mutations with long (JEV_LS and TBEV_LS) and short

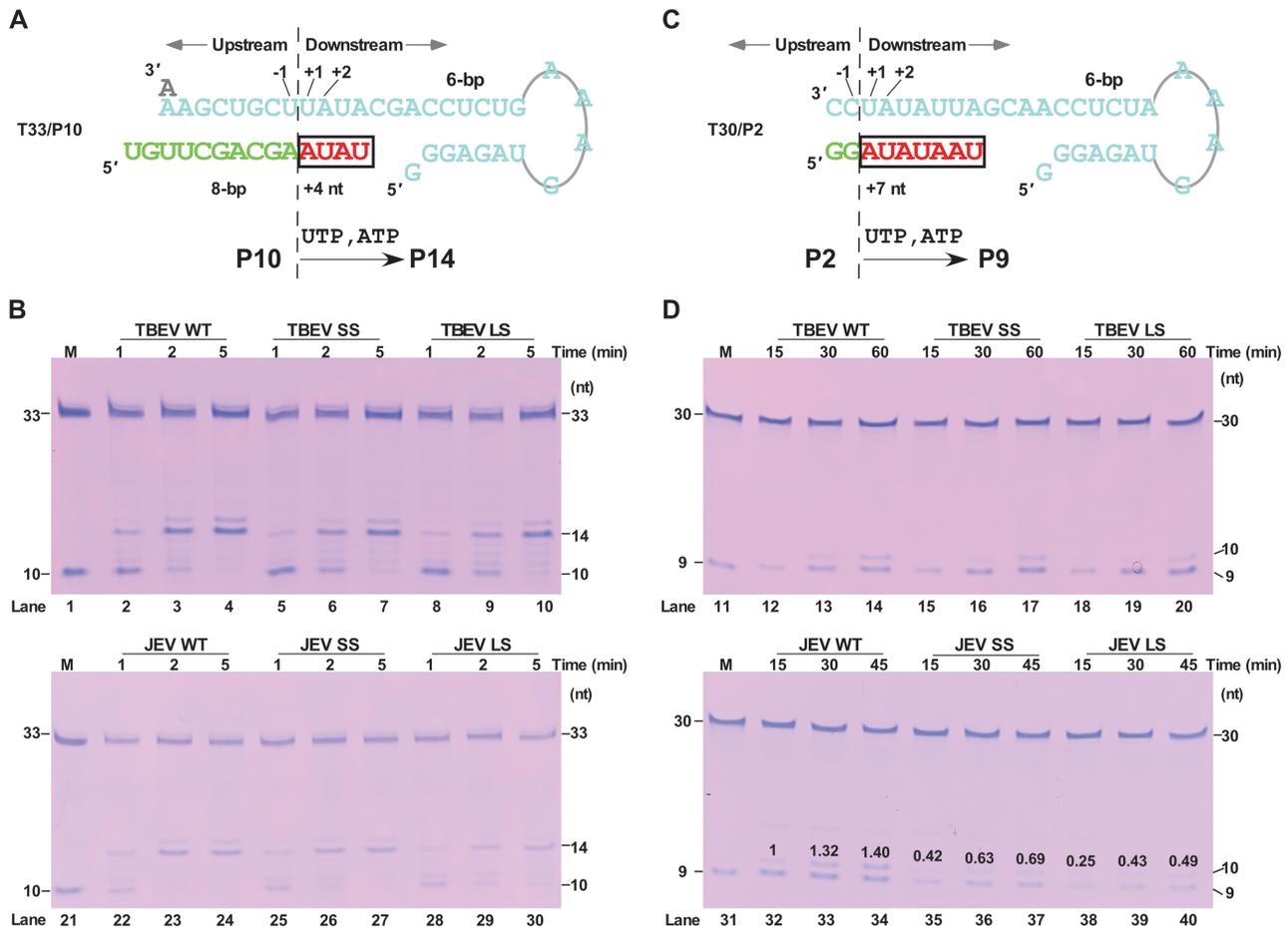


Figure 4. A comparison of *in vitro* polymerase activities of NS5s and its region B-C substitution mutants. (A) A diagram of construct T33/P10 used in the primer-dependent polymerase assays. The template strand (cyan) and the primer (green) form an 8-base pair (bp) duplex in the upstream region, and the downstream template folds back to form a stem-loop structure. (B) The primer-dependent polymerase activities of WT NS5 and its inter-virus region B-C substitution variants. The time-dependent conversion of the primer P10 to the 14–15-mer products was used to assess the polymerase activities upon ATP and UTP additions. (C) A diagram of construct T30/P2 used in the dinucleotide-driven polymerase assays. (D) The dinucleotide-driven polymerase activities of WT NS5 and its variants. The time-dependent conversion of the primer P2 to the 9–10-mer products was used to assess the polymerase activities upon ATP and UTP additions. The total amount of 9–10-mer intensity according to WT reaction at 15 min was used as the reference to estimate relative polymerase activities in lanes 33–40.

(JEV_SS and TBEV_SS) substitutions, respectively (Figure 3). Corresponding TBEV and JEV NS5 mutant constructs were then compared with WT enzymes in both primer-dependent and dinucleotide-driven RdRP assays to assess whether substitutions in region B-C affect polymerase catalytic activities. In previous studies characterizing flavivirus NS5 RdRP, DENV2 NS5 exhibited manganese ion (Mn^{2+})-independent activities *in vitro*, while JEV NS5 required Mn^{2+} in both primer-dependent and dinucleotide-driven assays (38,39). It turned out that TBEV NS5 also needs Mn^{2+} in our experimental settings. Using a 33-nucleotide (nt) RNA template (T33) and a 10-nt primer (P10) that can form an 8-base pair (bp) template-primer duplex, an A-U-A-U tetra-nucleotide was readily incorporated to make a 14-nt product (P14) within 5 min using WT TBEV and JEV enzymes, while a minor product 15-nt in length likely derived from a misincorporation event was also observed (Figure 4 A–B; Supplementary Figure S4, lanes 2–6). Region B-C substitutions in either TBEV or JEV NS5 did not apparently affect the product profiles and formation kinet-

ics (Figure 4B, compare lanes 2–4 with lanes 5–7 and 8–10 for TBEV constructs; compare lanes 22–24 with lanes 25–27 and 28–30 for JEV constructs). The dinucleotide-driven assays were typically used in various polymerase systems to mimic *de novo* synthesis (45–47), and a 30-nt template (T30) and a G–G dinucleotide with a 5'-phosphate (pGG) previously used in characterizing JEV, DENV and CSFV RdRPs were used in this study (Figure 4C). When ATP and UTP were used as the only NTP substrates, the formation of the 9 nt product for TBEV WT NS5 was slower than that of JEV WT enzyme (Figure 4D, compare lanes 12–14 with 32–34). The 10 nt product was observed in both enzyme systems, possibly derived from a G:U misincorporation event (Supplementary Figure S4, lanes 7–9). For substitutions using the TBEV RdRP backbone, no obvious effect on catalysis was observed (Figure 4D, compare lanes 12–14 with lanes 15–17 and 18–20). By contrast, substitutions in the JEV RdRP backbone resulted in slower formation of the 9–10 nt products (Figure 4D, compare lanes 32–34 with lanes 35–37 and 38–40). Taken together, the maintenance of RdRP

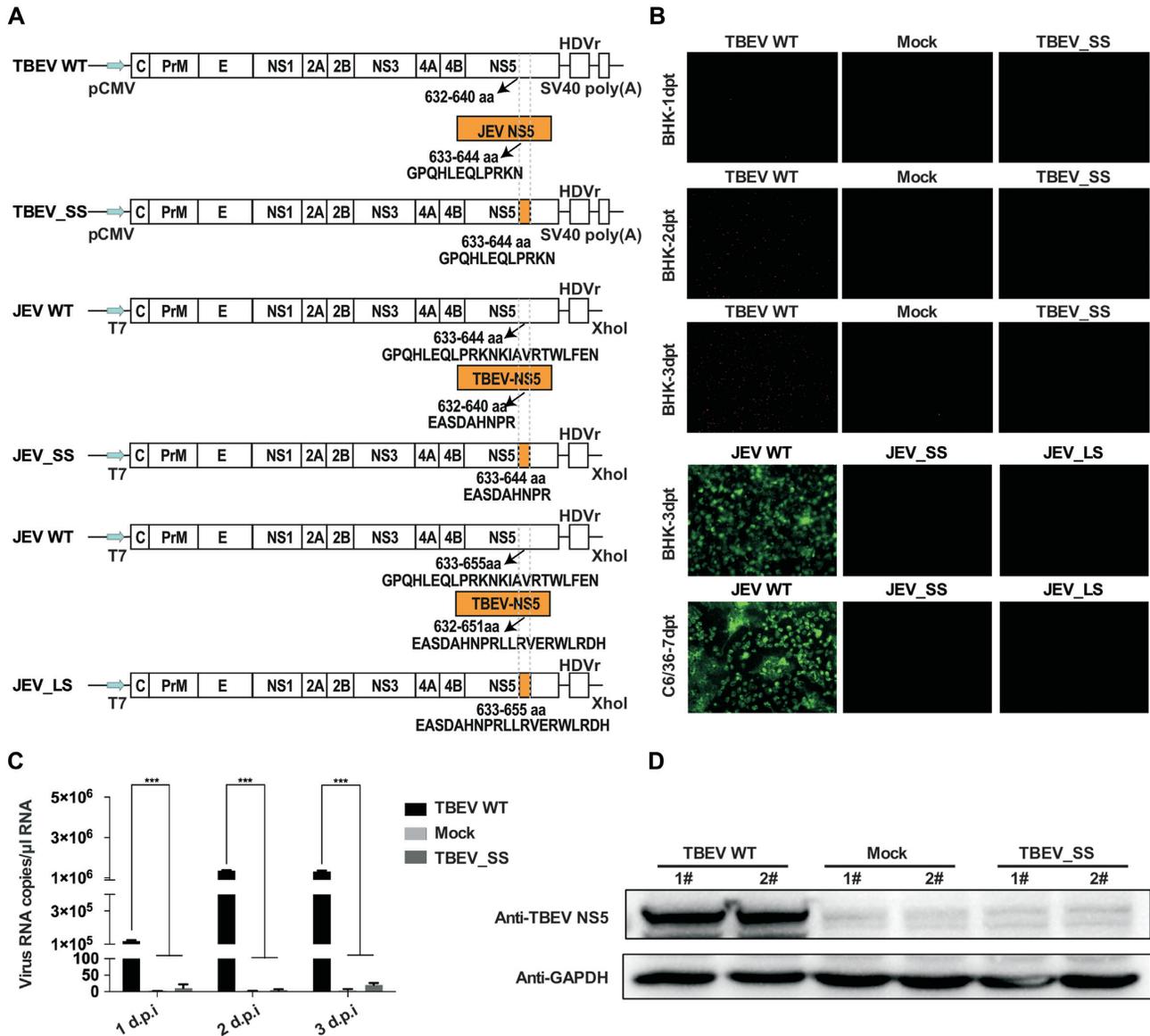


Figure 5. The inter-species region B-C substitution abolished virus proliferation. (A) Schematic diagram of WT infectious clones and the corresponding TBEV_SS, JEV_SS and JEV_LS mutant clones. For all three substitution mutants, the design of substitution is identical to those in Figure 3. (B) IFA analysis of WT and TBEV_SS, JEV_SS and JEV_LS mutants. Equal amounts of genomic RNAs of JEV WT, JEV_SS and JEV_LS mutants were transfected into BHK-21 and C6/36 cells, respectively. IFA was performed using JEV E (4G2) and TBEV NS5 antibodies to detect the viral protein synthesis at the indicated time points. (C) The viral RNA quantification of TBEV. The total mRNA of BHK-21 cell was extracted, and TBEV mRNA was detected at 1, 2 and 3 d.p.i. via RT-qPCR. (D) The western blotting assay detecting TBEV NS5 in the infected BHK-21 cells. GAPDH (glyceraldehyde-3-phosphate dehydrogenase) was used as a loading control.

activities after region B-C substitutions in both TBEV and JEV systems suggests that region B-C may not play essential roles in RdRP catalysis.

Inter-virus region B-C substitution abolishes TBEV and JEV proliferation but supports low-level RNA replication in a TBEV replicon

In order to assess the role of region B-C in flavivirus proliferation, SS substitution was introduced into a TBEV infectious clone, while the corresponding LS and SS substitutions were introduced into a JEV infectious clone (Figure 5A) (48–50). For the virus recovery of WT and mu-

tant TBEV, as described previously (49,50), equal amounts of infectious clone plasmids were transfected into BHK-21 cells to assess the effect brought by the substitution. Three days post transfection, the cells transfected with the WT infectious clone plasmids exhibited obvious cytopathic effect, and supernatant samples (for both WT and mutant) were then prepared and used to infect naïve BHK-21 cells (with a virus titer of 7×10^6 plaque forming units (pfu)/ml determined by a plaque assay) (Supplementary Figure S5). Viral NS5 protein production level was monitored by an immunofluorescence assay (IFA). IFA positive cells were not detected for 3 days post infection for the mutant clone, while obvious NS5 production was evident for the WT clone (Fig-

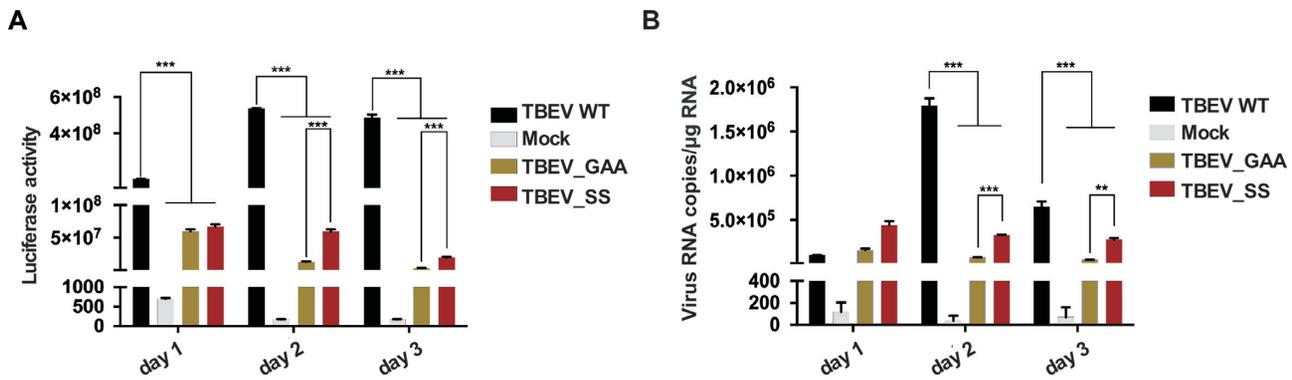


Figure 6. The inter-species region B-C substitution in a TBEV replicon supports low-level RNA replication. **(A)** Luciferase activity of the cells transiently transfected with TBEV replicon. The luciferase activities were measured at the indicated time points post transfection. Average results of three independent experiments are presented. Error bars represent standard deviations. **(B)** The quantification of TBEV genomic RNA. The total RNAs of BHK-21 cells were extracted, and TBEV genomic RNAs were detected at 1, 2 and 3 d.p.i via RT-qPCR.

ure 5B, top three rows). For the WT and mutant JEV clones, equal amounts of viral RNAs were transfected into BHK-21 or C6/36 cells to assess the effect brought by the mutations in both mammalian and mosquito host cells. Viral envelope (E) protein production level was monitored by IFA. In either host cell systems, IFA positive cells were not detected for both LS and SS substitution clones after transfection, while continuous production of viral E protein was observed for the WT clone (Figure 5B, bottom two rows). The TBEV IFA data were further validated by comparing the amount of viral RNA copies and antibody-based NS5 production levels, and the SS mutant virus failed to produce viral RNAs and NS5 proteins (Figure 5C–D). These data together suggest that region B-C plays critical roles in virus proliferation for both virus species. Although the LS/SS mutations may change the local RNA structure, it is less likely for these nucleotide mutations to abolish virus proliferation by disrupting conserved long-range interactions essential to flavivirus replication mainly involving conserved structural elements in the untranslated regions (51,52).

To further explore the functional relevance of region B-C, a TBEV replicon bearing the SS substitution and a luciferase reporter gene was constructed, and the luciferase activity and RNA copy numbers were used to assess the replicon-derived RNA replication in BHK-21 cells transfected with the replicon RNAs (Figure 6). Interestingly, the SS mutant replicon was able to support low level replication as indicated by the luciferase activities and RNA copy numbers if compared with WT replicon (about 1 order of magnitude lower than the WT), but it clearly replicated better than a mutant replicon bearing double D-to-A mutations at the flavivirus motif C catalytic sequence GDD (TBEV_GAA mutant, corresponding to residues 663–665 in TBEV NS5) (Figure 6). These data suggest that the SS substitution significantly reduced cellular level RNA replication but RdRP catalytic activities were not abolished. The non-detectable replication for the SS mutant virus and low-level replication of the SS-mutant replicon may suggest that the substitution mutation may also affect processes beyond those necessary to support replicon replication.

DISCUSSION

The host ranges of RNA viruses encompass almost all cellular forms of life. As the essential and most conserved enzyme encoded by RNA viruses, viral RdRPs have a long evolutionary history (53). On one hand, they need to retain a central catalytic module to perform the RNA-directed RNA synthesis. On the other hand, host adaptation-related factors have driven them to exhibit high-level diversity in overall architecture and regulatory mechanism (44,54). Based on known RdRP structures, the size of RdRP central catalytic module is about 450–600 residues, forming the aforementioned encircled right hand structure with seven catalytic motifs (A–G) distributed around the active site in a highly similar manner (55–57). The 461-residue poliovirus (PV) 3D^{pol} and the 591-residue hepatitis C virus (HCV) may represent the most compact forms of primer-dependent and *de novo* RdRPs, respectively, with the latter having a bulkier thumb harboring structural element to facilitate *de novo* initiation (58,59). By contrast, the 2000–4000-residue *Bunyavirales* L proteins represent the other extreme with multiple functional regions, including the cap-snatching-related cap binding and endonuclease modules, integrated with the RdRP catalytic module (60,61). The RdRPs encoded by some of the positive-strand RNA viruses tend to have N-terminal functional regions fused to RdRP catalytic module (44). The pestivirus NS5B contains a 90-residue N-terminal domain (NTD) that regulates the fidelity of RdRP through intra-molecular interactions with the palm domain (62,63). The flavivirus NS5 harbors a 260-residue MTase that interacts with the RdRP through at least two modes of intra-molecular interactions (17,18). While both interaction modes revealed by crystallography are critical for virus proliferation, only one of them contributes to RdRP catalysis by improving NTP binding at the initiation stage of the synthesis (39,64). The coronavirus RdRP nsp12 contains a 250-residue NiRAN (nidovirus RdRP-associated nucleotidyltransferase) domain and a 120-residue interface domain that fold with the RdRP palm and fingers. However, the function of these two domains remains to be clarified (65–68).

While the global structures of viral RdRPs are quite diverse mainly due to the integration of other functional re-

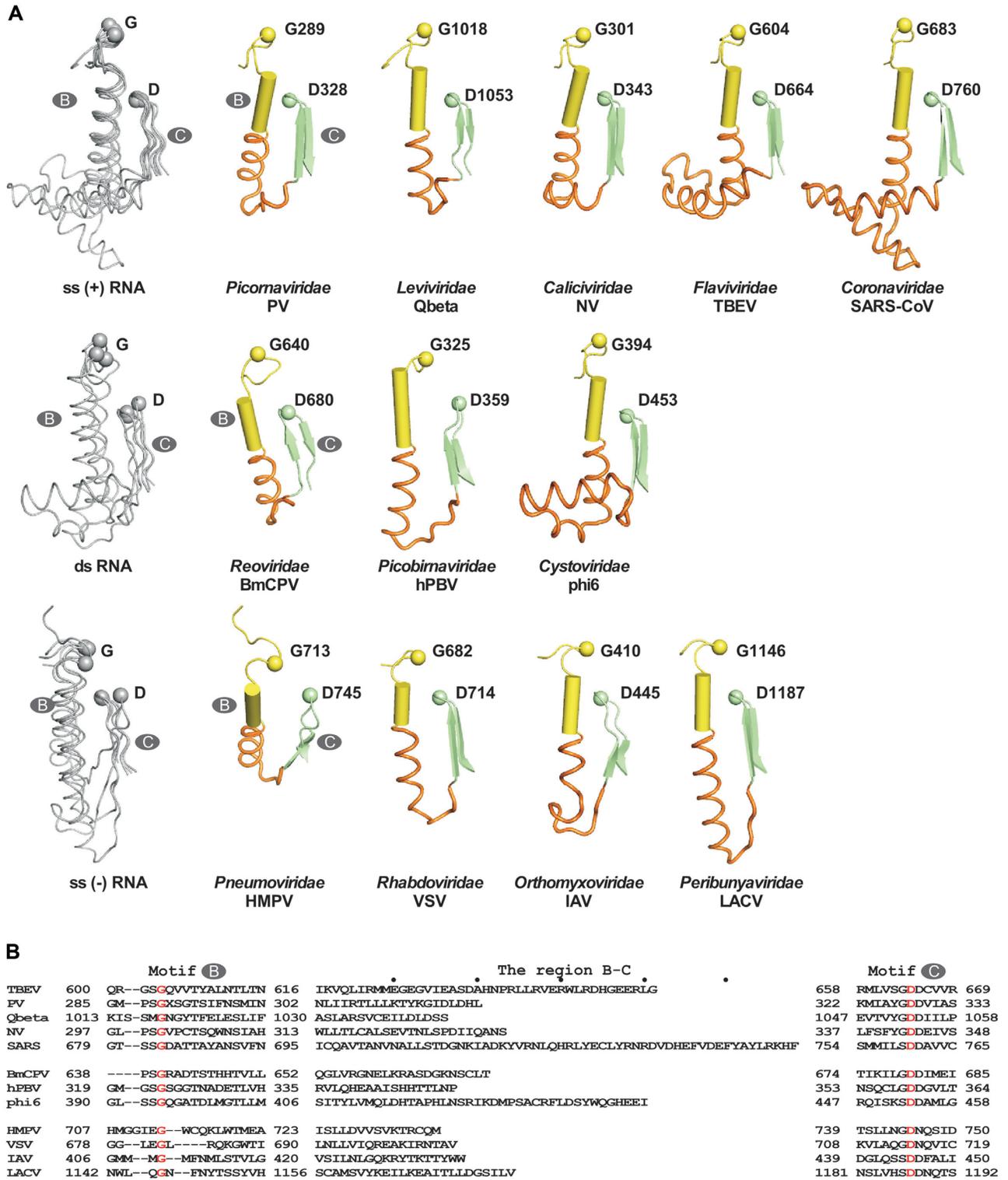


Figure 7. A diversity survey of region B-C region in viral RdRPs. (A) Structural comparison of the motifs B-C region in RNA virus polymerases. Each structure represents an RNA virus family. Coloring scheme: motif B in yellow, region B-C in orange, motif C in light green. Two absolutely conserved residues (the G in the SG sequence of motif B and the first D in XGDD of motif C) are shown as spheres. The α -helix in motif B is shown as a cylinder. ss (+) RNA, ss (-) RNA, ds RNA are used as abbreviations for positive-strand, negative-strand, and double-strand RNA viruses, respectively. (B) Structure-based sequence alignment of RdRP motifs B and C, with region B-C sequences listed (not aligned) for comparison. Two absolutely conserved residues are highlighted in red.

gions and co-folding with other viral proteins (68,69), the RdRP catalytic module structure is relatively conserved. In particular, the spatial arrangement of motif A-G is consistent for all RdRPs despite low sequence conservation level in the scope of all RNA viruses (54,70). Typically, the seven motifs appear in the order of G-F-A-B-C-D-E in RdRP primary structure, except for the *Permutatetetraviridae* and *Birnaviridae* RdRPs that have a different order of G-F-C-A-B-D-E (57,71). Among them, motifs A, B, C, and F directly interact with the NTP substrate and contain highly conserved residues (42,72). In contrast, motifs D, E and G located at the periphery of the active site mainly play structural roles and are less conserved in sequence (70,73). Hence the residue spacing among motifs F-A-B-C can be readily surveyed in majority of RdRPs by locating the highly conserved residues. While the residue spacing for motifs F/A and A/B is relatively consistent, and is about 40–60 residues for both regions, the spacing between motifs B and C varies to a much greater extent (44,55). The shortest spacing is smaller than 20-residue as seen in RdRPs from *Tymovirales*, *Togaviridae*, *Alvernaviridae* etc., while the longest is larger than 70-residue in *Mononiviridae* (belonging to *Nidovirales*) RdRPs in a survey of RdRPs from positive-strand RNA viruses (44).

The observations of large distance variation between RdRP motifs B and C echo the aforementioned structure and length variations among *flavivirus* and *Flaviviridae* RdRP region B-C, and invoke us to further survey the corresponding region of other representative RdRP structures. For other positive-strand RNA viruses, region B-C lengths of RdRPs from *Picornaviridae*, *Caliciviridae* and *Leviviridae* are relatively short, while those from *Coronaviridae* are the longest and forming a helix bundle (Figure 7A, top panel; Figure 7B). As revealed in the SARS-CoV-2 RdRpnsp7-nsp8 complex structure, part of region B-C interacts with the NiRAN domain hairpin, while the rest of it is solvent accessible (65). For double stranded RNA viruses, the *Reoviridae* and *Picobirnaviridae* RdRP region B-Cs are short and comparable to the *Picornaviridae* counterparts, while that from the *Cystoviridae* is longer and similar to the flavivirus region B-C both in length and structure (Figure 7A, middle panel; Figure 7B). For negative-strand RNA viruses including *Pneumoviridae*, *Rhabdoviridae*, *Orthomyxoviridae*, and *Peribunyaviridae*, the lengths of RdRP region B-C are relatively short, but the spatial placement (orientation) of this region relative to RdRP active site varies to some extent (Figure 7A, bottom panel; Figure 7B).

Taken together, the length and structure diversity of region B-C and its solvent accessible feature suggest that it may have become a regulatory hotspot in RdRP evolution history. Inter-species substitution retained RdRP catalytic activities but did not support virus proliferation further support the possibility of its involvement in other viral/host events. Through identifying relatively subtle differences of region B-Cs between the tick-borne and mosquito-borne flavivirus RdRPs and generalizing its high-level of diversity in the context of viral RdRPs, the current study pointed out this probable host adaptation-related region within the most conserved enzyme module of RNA viruses. However, the specific role of this region and the precise mechanism underneath, either in flavivirus or in others, require further

investigations including discoveries of region B-C interactions with viral and host elements.

DATA AVAILABILITY

The atomic coordinates and structure factors for the reported crystal structures of the TBEV NS5 MTase and RdRP have been deposited in the Protein Data Bank under accession numbers 7D6M and 7D6N, respectively.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Wuhan National Biosafety Laboratory for access to BSL-3 laboratories, Dr Qinxue Hu for providing the cloning material for the TBEV NS5 gene, Dr Gengfu Xiao and Dr Ke Peng for helpful discussion, Dr Penghui Li, Yuanjiu Miao, Dr Jiqin Wu, Buyu Zhang, and Liu Deng for laboratory assistance, synchrotron SSRF (beamline BL17U Shanghai, China) for access to beamline, and The Core Facility and Technical Support, Wuhan Institute of Virology, for access to instruments.

Author contributions: P.G., H.W., B.Z. and Z.Z. conceived the research; J.Y. performed the crystallographic and biochemical experiments; W.Y. and X.-D.L. performed the TBEV- and JEV-based virological experiments, respectively; X.J. and J.Y. collected the X-ray diffraction data; J.Y., X.J. and P.G. analyzed the crystallographic and biochemical data; W.Y., H.W. and Z.Z. analyzed the TBEV virological data; X.-D.L. and B.Z. analyzed the JEV virological data; J.Y., X.J. and P.G. assembled the initial manuscript and all authors contributed to the writing of the paper.

FUNDING

National Key Research and Development Program of China [2018YFD0500100]; National Natural Science Foundation of China [32070185, 32041007, 31670154]; Chinese Academy of Sciences Funds: the Advanced Customer Cultivation Project of Wuhan National Biosafety Laboratory [2018ACCP-MS06]; ‘One-Three-Five’ Strategic Programs, Wuhan Institute of Virology [Y605191SA1]. Funding for open access charge: National Key Research and Development Program of China [2016YFD0500100]; National Natural Science Foundation of China [32070185, 32041007, 31670154].

Conflict of interest statement. None declared.

REFERENCES

- Gaunt, M.W., Sall, A.A., Lamballerie, X., Falconar, A.K.I., Dzhivanian, T.I. and Gould, E.A. (2001) Phylogenetic relationships of flaviviruses correlate with their epidemiology, disease association and biogeography. *J. Gen. Virol.*, **82**, 1867–1876.
- Kuno, G., Chang, G.J., Tsuchiya, K.R., Karabatsos, N. and Cropp, C.B. (1998) Phylogeny of the genus *Flavivirus*. *J. Virol.*, **72**, 73–83.
- Billoir, F., de Chesse, R., Tolou, H., de Micco, P., Gould, E.A. and de Lamballerie, X. (2000) Phylogeny of the genus *flavivirus* using complete coding sequences of arthropod-borne viruses and viruses with no known vector. *J. Gen. Virol.*, **81**, 2339.

4. Holbrook, M.R. (2017) Historical perspectives on flavivirus research. *Viruses*, **9**, 97.
5. Gubler, D.J., Kuno, G. and Markoff, L. (2007) In: *Fields Virology*, 5th edn. *Flaviviruses*.
6. Pettersson, J.H. and Fik-Palacios, O. (2014) Dating the origin of the genus *Flavivirus* in the light of Beringian biogeography. *J. Gen. Virol.*, **95**, 1969–1982.
7. Weaver, S.C. and Barrett, A.D. (2004) Transmission cycles, host range, evolution and emergence of arboviral disease. *Nat. Rev. Microbiol.*, **2**, 789–801.
8. Lasala, P.R. and Holbrook, M. (2010) Tick-borne flaviviruses. *Clin. Lab. Med.*, **30**, 221–235.
9. Grard, G., Moureau, G., Charrel, R.N., Lemasson, J.J., Gonzalez, J.P., Gallian, P., Gritsun, T.S., Holmes, E.C., Gould, E.A. and de Lamballerie, X. (2007) Genetic characterization of tick-borne flaviviruses: new insights into evolution, pathogenetic determinants and taxonomy. *Virology*, **361**, 80–92.
10. Gritsun, T.S., Nuttall, P.A. and Gould, E.A. (2003) Tick-borne flaviviruses. *Adv. Virus Res.*, **61**, 317–371.
11. Blitvich, B.J. and Firth, A.E. (2017) A review of flaviviruses that have no known arthropod vector. *Viruses*, **9**, 154.
12. Blitvich, B.J. and Firth, A.E. (2015) Insect-specific flaviviruses: a systematic review of their discovery, host range, mode of transmission, superinfection exclusion potential and genomic organization. *Viruses*, **7**, 1927–1959.
13. Cook, S., Moureau, G., Kitchen, A., Gould, E.A., de Lamballerie, X., Holmes, E.C. and Harbach, R.E. (2012) Molecular evolution of the insect-specific flaviviruses. *J. Gen. Virol.*, **93**, 223–234.
14. Kuno, G. (2007) Host range specificity of flaviviruses: correlation with in vitro replication. *J. Med. Entomol.*, **44**, 93–101.
15. Proutski, V., Gould, E.A. and Holmes, E.C. (1997) Secondary structure of the 3' untranslated region of flaviviruses: similarities and differences. *Nucleic Acids Res.*, **25**, 1194–1202.
16. Wallner, G., Mandl, C.W., Kunz, C. and Heinz, F.X. (1995) The flavivirus 3'-noncoding region: extensive size heterogeneity independent of evolutionary relationships among strains of tick-borne encephalitis virus. *Virology*, **213**, 169–178.
17. Lu, G. and Gong, P. (2013) Crystal structure of the Full-Length Japanese encephalitis virus NS5 reveals a conserved Methyltransferase-Polymerase interface. *PLoS Pathog.*, **9**, e1003549.
18. Zhao, Y., Soh, T.S., Zheng, J., Chan, K.W., Phoo, W.W., Lee, C.C., Tay, M.Y., Swaminathan, K., Cornvik, T.C., Lim, S.P. et al. (2015) A crystal structure of the Dengue virus NS5 protein reveals a novel inter-domain interface essential for protein flexibility and virus replication. *PLoS Pathog.*, **11**, e1004682.
19. Egloff, M.P., Benarroch, D., Selisko, B., Romette, J.L. and Canard, B. (2002) An RNA cap (nucleoside-2'-O)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization. *EMBO J.*, **21**, 2757–2768.
20. Malet, H., Egloff, M.P., Selisko, B., Butcher, R.E., Wright, P.J., Roberts, M., Gruez, A., Sulzenbacher, G., Vornrhein, C., Bricogne, G. et al. (2007) Crystal structure of the RNA polymerase domain of the West Nile virus non-structural protein 5. *J. Biol. Chem.*, **282**, 10678–10689.
21. Mastrangelo, E., Bollati, M., Milani, M., Selisko, B., Peyrane, F., Canard, B., Grard, G., de Lamballerie, X. and Bolognesi, M. (2007) Structural bases for substrate recognition and activity in Meaban virus nucleoside-2'-O-methyltransferase. *Protein Sci.*, **16**, 1133–1145.
22. Shi, M., Lin, X.D., Vasilakis, N., Tian, J.H., Li, C.X., Chen, L.J., Eastwood, G., Diao, X.N., Chen, M.H., Chen, X. et al. (2016) Divergent viruses discovered in arthropods and vertebrates revise the evolutionary history of the flaviviridae and related viruses. *J. Virol.*, **90**, 659–669.
23. Temmam, S., Bigot, T., Chretien, D., Gondard, M., Perot, P., Pommelet, V., Dufour, E., Petres, S., Devillers, E., Hoem, T. et al. (2019) Insights into the host range, genetic diversity, and geographical distribution of Jingmenviruses. *mSphere*, **4**, e00645-19.
24. Jia, N., Liu, H.B., Ni, X.B., Bell-Sakyi, L., Zheng, Y.C., Song, J.L., Li, J., Jiang, B.G., Wang, Q., Sun, Y. et al. (2019) Emergence of human infection with Jingmen tick virus in China: A retrospective study. *EBioMedicine*, **43**, 317–324.
25. Taniguchi, S. (2019) Detection of Jingmen tick virus in human patient specimens: emergence of a new tick-borne human disease? *EBioMedicine*, **43**, 18–19.
26. Qin, X.C., Shi, M., Tian, J.H., Lin, X.D., Gao, D.Y., He, J.R., Wang, J.B., Li, C.X., Kang, Y.J., Yu, B. et al. (2014) A tick-borne segmented RNA virus contains genome segments derived from unsegmented viral ancestors. *PNAS*, **111**, 6744–6749.
27. Chiu, J., March, P.E., Lee, R. and Tillett, D. (2004) Site-directed, ligase-independent mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h. *Nucleic Acids Res.*, **32**, e174.
28. Chiu, J., Tillett, D., Dawes, I.W. and March, P.E. (2008) Site-directed, ligase-independent mutagenesis (SLIM) for highly efficient mutagenesis of plasmids greater than 8kb. *J. Microbiol. Methods*, **73**, 195–198.
29. Papworth, C.B., Bauer, J.C., Braman, J. and Wright, D.A. (1996) Site-directed mutagenesis in one day with >80% efficiency. *Strategies*, **9**, 3–4.
30. Gohara, D.W., Ha, C.S., Kumar, S., Ghosh, B., Arnold, J.J., Wisniewski, T.J. and Cameron, C.E. (1999) Production of "authentic" poliovirus RNA-dependent RNA polymerase (3D(pol)) by ubiquitin-protease-mediated cleavage in *Escherichia coli*. *Protein Expr. Purif.*, **17**, 128–138.
31. Otwinowski, Z. and Minor, W. (1997) In: *Processing of X-ray Diffraction Data Collected in Oscillation Mode*. Academic Press, NY.
32. Kabsch, W. (2010) Xds. *Acta Crystallogr. D, Biol. Crystallogr.*, **66**, 125–132.
33. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D, Biol. Crystallogr.*, **66**, 486–501.
34. Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W. et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D, Biol. Crystallogr.*, **66**, 213–221.
35. Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S. et al. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D, Biol. Crystallogr.*, **54**, 905–921.
36. Theobald, D.L. and Wuttke, D.S. (2006) THESEUS: maximum likelihood superpositioning and analysis of macromolecular structures. *Bioinformatics*, **22**, 2171–2172.
37. Batey, R.T. and Kieft, J.S. (2007) Improved native affinity purification of RNA. *RNA*, **13**, 1384–1389.
38. Wu, J., Lu, G., Zhang, B. and Gong, P. (2015) Perturbation in the conserved methyltransferase-polymerase interface of flavivirus NS5 differentially affects polymerase initiation and elongation. *J. Virol.*, **89**, 249–261.
39. Wu, J., Ye, H.Q., Zhang, Q.Y., Lu, G., Zhang, B. and Gong, P. (2020) A conformation-based intra-molecular initiation factor identified in the flavivirus RNA-dependent RNA polymerase. *PLoS Pathog.*, **16**, e1008484.
40. Li, X.D., Li, X.F., Ye, H.Q., Deng, C.L., Ye, Q., Shan, C., Shang, B.D., Xu, L.L., Li, S.H., Cao, S. et al. (2013) Recovery of a chemically synthesized Japanese encephalitis virus revealed two critical adaptive mutations in NS2B and NS4A. *J. Gen. Virol.*, **95**, 806–815.
41. Yap, T.L., Xu, T., Chen, Y.L., Malet, H., Egloff, M.P., Canard, B., Vasudevan, S.G. and Lescar, J. (2007) Crystal structure of the dengue virus RNA-dependent RNA polymerase catalytic domain at 1.85-angstrom resolution. *J. Virol.*, **81**, 4753–4765.
42. Gong, P. and Peersen, O.B. (2010) Structural basis for active site closure by the poliovirus RNA-dependent RNA polymerase. *PNAS*, **107**, 22505–22510.
43. Zamyatkin, D.F., Parra, F., Alonso, J.M., Harki, D.A., Peterson, B.R., Grochulski, P. and Ng, K.K. (2008) Structural insights into mechanisms of catalysis and inhibition in Norwalk virus polymerase. *J. Biol. Chem.*, **283**, 7705–7712.
44. Jia, H. and Gong, P. (2019) A Structure-Function Diversity Survey of the RNA-Dependent RNA Polymerases From the Positive-Strand RNA Viruses. *Front. Microbiol.*, **10**, 1945.
45. Samuels, M., Fire, A. and Sharp, P.A. (1984) Dinucleotide priming of transcription mediated by RNA polymerase II. *J. Biol. Chem.*, **259**, 2517–2525.
46. Jin, Z., Leveque, V., Ma, H., Johnson, K.A. and Klumpp, K. (2012) Assembly, purification, and pre-steady-state kinetic analysis of active RNA-dependent RNA polymerase elongation complex. *J. Biol. Chem.*, **287**, 10674–10683.

47. Honda, A., Mizumoto, K. and Ishihama, A. (1986) RNA polymerase of influenza virus. Dinucleotide-primed initiation of transcription at specific positions on viral RNA. *J. Biol. Chem.*, **261**, 5987–5991.
48. Li, X.D., Li, X.F., Ye, H.Q., Deng, C.L., Ye, Q., Shan, C., Shang, B.D., Xu, L.L., Li, S.H., Cao, S.B. *et al.* (2014) Recovery of a chemically synthesized Japanese encephalitis virus reveals two critical adaptive mutations in NS2B and NS4A. *J. Gen. Virol.*, **95**, 806–815.
49. Li, P., Ke, X., Wang, T., Tan, Z., Luo, D., Miao, Y., Sun, J., Zhang, Y., Liu, Y., Hu, Q. *et al.* (2018) Zika virus attenuation by codon pair deoptimization induces sterilizing immunity in mouse models. *J. Virol.*, **92**, e00701-18.
50. Wang, T., Li, P., Zhang, Y., Liu, Y., Tan, Z., Sun, J., Ke, X., Miao, Y., Luo, D., Hu, Q. *et al.* (2020) In vivo imaging of Zika virus reveals dynamics of viral invasion in immune-sheltered tissues and vertical propagation during pregnancy. *Theranostics*, **10**, 6430–6447.
51. Dong, H., Zhang, B. and Shi, P.Y. (2008) Terminal structures of West Nile virus genomic RNA and their interactions with viral NS5 protein. *Virology*, **381**, 123–135.
52. Gebhard, L.G., Filomatori, C.V. and Gamarnik, A.V. (2011) Functional RNA elements in the dengue virus genome. *Viruses*, **3**, 1739–1756.
53. Wolf, Y.I., Kazlauskas, D., Iranzo, J., Lucia-Sanz, A., Kuhn, J.H., Krupovic, M., Dolja, V.V. and Koonin, E.V. (2018) Origins and evolution of the global RNA virome. *mBio*, **9**, e02329-18.
54. te Velthuis, A.J. (2014) Common and unique features of viral RNA-dependent polymerases. *Cell. Mol. Life Sci.*, **71**, 4403–4420.
55. Poch, O., Sauvaget, L., Delarue, M. and Tordo, N. (1989) Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J.*, **8**, 3867–3874.
56. Bruenn, J.A. (2003) A structural and primary sequence comparison of the viral RNA-dependent RNA polymerases. *Nucleic Acids Res.*, **31**, 1821–1829.
57. Gorbalenya, A.E., Pringle, F.M., Zeddam, J.L., Luke, B.T., Cameron, C.E., Kalmakoff, J., Hanzlik, T.N., Gordon, K.H. and Ward, V.K. (2002) The palm subdomain-based active site is internally permuted in viral RNA-dependent RNA polymerases of an ancient lineage. *J. Mol. Biol.*, **324**, 47–62.
58. Thompson, A.A. and Peersen, O.B. (2004) Structural basis for proteolysis-dependent activation of the poliovirus RNA-dependent RNA polymerase. *EMBO J.*, **23**, 3462–3471.
59. Lesburg, C.A., Cable, M.B., Ferrari, E., Hong, Z., Mannarino, A.F. and Weber, P.C. (1999) Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site. *Nat. Struct. Biol.*, **6**, 937–943.
60. Gerlach, P., Malet, H., Cusack, S. and Reguera, J. (2015) Structural insights into bunyavirus replication and its regulation by the vRNA promoter. *Cell*, **161**, 1267–1279.
61. Kinsella, E., Martin, S.G., Grolla, A., Czub, M., Feldmann, H. and Flick, R. (2004) Sequence determination of the Crimean-Congo hemorrhagic fever virus L segment. *Virology*, **321**, 23–28.
62. Liu, W., Shi, X. and Gong, P. (2018) A unique intra-molecular fidelity-modulating mechanism identified in a viral RNA-dependent RNA polymerase. *Nucleic Acids Res.*, **46**, 10840–10854.
63. Li, W., Wu, B., Soca, W.A. and An, L. (2018) Crystal structure of classical swine fever virus NS5B reveals a novel N-Terminal domain. *J. Virol.*, **92**, e00324–18.
64. Li, X.D., Shan, C., Deng, C.L., Ye, H.Q., Shi, P.Y., Yuan, Z.M., Gong, P. and Zhang, B. (2014) The interface between methyltransferase and polymerase of NS5 is essential for flavivirus replication. *PLoS Negl. Trop. Dis.*, **8**, e2891.
65. Gao, Y., Yan, L., Huang, Y., Liu, F., Zhao, Y., Cao, L., Wang, T., Sun, Q., Ming, Z., Zhang, L. *et al.* (2020) Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science*, **368**, 779–782.
66. Lehmann, K.C., Gulyaeva, A., Zevenhoven-Dobbe, J.C., Janssen, G.M., Ruben, M., Overkleeft, H.S., van Veen, P.A., Samborskiy, D.V., Kravchenko, A.A., Leontovich, A.M. *et al.* (2015) Discovery of an essential nucleotidylating activity associated with a newly delineated conserved domain in the RNA polymerase-containing protein of all nidoviruses. *Nucleic Acids Res.*, **43**, 8416–8434.
67. Wang, Q., Wu, J., Wang, H., Gao, Y., Liu, Q., Mu, A., Ji, W., Yan, L., Zhu, Y., Zhu, C. *et al.* (2020) Structural basis for RNA replication by the SARS-CoV-2 polymerase. *Cell*, **182**, 417–428.
68. Kirchdoerfer, R.N. and Ward, A.B. (2019) Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat. Commun.*, **10**, 2342.
69. Pflug, A., Guilligay, D., Reich, S. and Cusack, S. (2014) Structure of influenza A polymerase bound to the viral RNA promoter. *Nature*, **516**, 355–360.
70. Wu, J., Liu, W. and Gong, P. (2015) A structural overview of RNA-dependent RNA polymerases from the flaviviridae family. *Int. J. Mol. Sci.*, **16**, 12943–12957.
71. Pan, J., Vakharia, V.N. and Tao, Y.J. (2007) The structure of a birnavirus polymerase reveals a distinct active site topology. *PNAS*, **104**, 7385–7390.
72. Appleby, T.C., Perry, J.K., Murakami, E., Barauskas, O., Feng, J., Cho, A., Fox, D. 3rd, Wetmore, D.R., McGrath, M.E., Ray, A.S. *et al.* (2015) Viral replication. Structural basis for RNA replication by the hepatitis C virus polymerase. *Science*, **347**, 771–775.
73. Wang, M., Li, R., Shu, B., Jing, X., Ye, H.Q. and Gong, P. (2020) Stringent control of the RNA-dependent RNA polymerase translocation revealed by multiple intermediate structures. *Nat. Commun.*, **11**, 2605.