A re-emerging arbovirus in livestock in China: Early genomic surveillance and phylogeographic analysis

Shuo Su[1,*]

1. Jiangsu Engineering Laboratory of Animal Immunology, Institute of Immunology, College of Veterinary Medicine, Nanjing Agricultural University, Nanjing, China

*Corresponding author: Shuo Su (shuosu@njau.edu.cn)

Runing Title: Genomic surveillance and phylogeographic analysis of GETV

Abstract

Viruses in livestock represent a great risk to public health due to the close contact of their hosts with humans and the potential long-range transportation though animal trade. Here, we show how to predict outbreaks of potential zoonotic pathogens and use spatially-explicit phylogeographic and phylodynamic approaches to provide estimates of key epidemiological parameters. First, we use metagenomic next generation sequencing (mNGS) to identify a Getah virus (GETV) as the pathogen responsible for a re-emerging swine disease in China. The GETV isolate is able to replicate in a variety of cell lines including human cells and shows high pathogenicity in a mouse model suggesting a potential public health risk. We obtained 16 complete genomes and 78 E2 gene sequences from viral strains collected within China from 2017 to 2021 through large-scale surveillance among livestock, pets and mosquitoes. Moreover, phylogenetic analysis reveals that three major GETV lineages are responsible for the current epidemic in livestock in China. We identify three potential positively selected sites and mutations of interest in E2, which may impact the transmissibility and pathogenicity of the virus. We then reconstruct the evolutionary and dispersal history of the virus and test the impact of several environmental factors on the viral genetic diversity through time and on the dispersal dynamic of viral lineages. Of note, we identify temporal variation in livestock meat consumption as a main predictor of viral genetic diversity through time. Finally, phylogeographic analyses indicate that GETV preferentially circulates within areas associated with relatively higher mean annual temperature and pig population density. Our results highlight the importance of continuous surveillance of GETV among livestock in southern Chinese regions associated with relatively high temperatures, and the need to control mosquitoes in the livestock farms. Our analyses of GETV also provide a baseline for future studies of the molecular epidemiology and early warning of emerging arboviruses in China.

Key words: Predict outbreaks of zoonotic pathogens; Getah virus; Genomic

surveillance; Next generation sequencing; Phylogeography

Introduction

Approximately 18% of emerging infectious diseases (EIDs) that affect humans originate from wild animals or livestock [1-4]. In many of these host reservoir species, emerging viruses appear to be well adapted, with little or no evidence of clinical disease. However, when these viruses spill over into humans, the effects can sometimes be devastating [5, 6]. Because livestock can often act as a conduit for pathogen spillover into susceptible human populations, research on emerging viral diseases is focused on livestock infections that often occur due to contact with wild animals [7]. For example, the swine industry couples high-density farming with international trade, thus generating a high risk of emerging virus transmission and the potential for global spread [8]. Moreover, the swine industry will increasingly represent such risk due to its constant growth to fulfill a high demand for pork. A disease outbreak caused by a new or emerging virus may incur substantial economic burden and also endanger human health due to close human contact with pigs. Pigs have also been shown to be a significant source of zoonotic viruses such as Nipah virus in Malaysia [9] or influenza A (H1N1) virus (IAV) that caused the “swine-origin influenza” pandemic [10]. The 2009 A/H1N1 influenza pandemic in Mexico arose from viruses circulating in pigs in central-west Mexico for more than a decade. The virus originated from Eurasia (the landmass containing Europe and Asia) owing to an expansion of influenza A virus diversity in swine resulting from long-distance live swine trade [11]. The importance of pigs as a source of emerging viruses has recently been illustrated by four cases of human acute encephalitis that were associated with a variant strain of pseudorabies virus, with all the patients having had close occupational contact with pigs [12]. An efficient approach to detect both known and unexpected novel viruses in a single test would therefore be crucial for emerging viral outbreak identification and management in swine worldwide.

Our ability to predict outbreaks of potential zoonotic pathogens requires an understanding of their ecology and evolution in reservoir hosts. Metagenomic next-

generation sequencing (mNGS) technologies are particularly suitable for identifying viral etiologies. The analysis of the virome, often referred to as the assemblage of viruses in metagenomic studies, can detect known and novel viruses in environmental, human, or animal samples [13, 14]. mNGS is very well-suited for early diagnosis and monitoring of novel porcine viral diseases due to its high accuracy, fast response (generating large data in a short amount of time) and high sensitivity [15]. When coupling the pathogen genomes assembled from mNGS with phylodynamic analyses, researchers are able to achieve a comprehensive understanding of the spatiotemporal patterns of spread and how these patterns have been shaped by external factors for zoonotic pathogens of epidemiological importance. In particular, relatively recent methodological developments allow for phylodynamic and phylogeographic approaches to test epidemiological hypotheses. For instance, the skygrid coalescent model [16] has been extended to allow for testing associations between the evolution of the virus effective population size through time and time series covariates [17]. Furthermore, discrete [18] or continuous [19, 20] phylogeographic reconstructions can be exploited to examine how covariates may explain the dispersal process of viral lineages [21-23]. The combination of mNGS technology and state-of-the-art analytical methods equips researchers with rapid identification of important emerging viruses and insight into the important epidemiological and environmental factors that shape its evolution, spread and hazards, providing a basis for its control and prevention.

Since May 2019, more than 1500 piglets died suddenly in four intensive pig farms in Guangxi, Henan, Hubei and Shandong Provinces, China, but the causal pathogen could not be identified by conventional diagnostic techniques. Eventually, metagenome sequencing of tissue samples from diseased piglets in our laboratory linked those cases to porcine Getah virus. Getah virus (GETV), an arbovirus and a member of the genus *Alphavirus*, can cause disease in domestic animals, including fever, rashes, edema of the hindlegs, and lymph node enlargement in horses, while infected piglets exhibit depression, tremors, hind limb paralysis, diarrhea, high mortality, and abortions [24-26]. GETV has a linear, positive-sense single stranded RNA genome encoding nine

viral proteins (nsP1-nsP4, E1-E3, C, and 6K). E2 is the main glycoprotein that binds to host cell receptors when initiating cell entry, whereas the E1 glycoprotein is required for pH-triggered membrane fusion within acidified endosomes [27]. Previous research shows that GETV gradually evolved within a relative broad host range [28]. Infections reported in mosquitoes, swine, cattle, horses, and blue foxes [29-32] suggest a wide distribution of susceptible animals in China. In addition, GETV neutralizing antibodies were detected in goats, cattle, horses, pigs, other animals [33], and humans [34], suggesting a potential public health risk. In the past 50 years, numerous *Alphavirus* re-emergences such as Chikungunya virus (CHIKV) have been documented in Africa and Asia, with irregular intervals of 2–20 years between outbreaks. These outbreaks led to many human infections, and even caused severe symptoms or death, as illustrated by Venezuelan equine encephalitis [35]. Therefore, a sudden outbreak of *Alphaviruses* not only poses a threat to the breeding industry, but also a potential threat to public health [36, 37].

In this study, we characterize and analyze the GETV outbreak occurring in the Chinese swine population through next-generation sequencing, an outbreak associated with a considerable impact on public, veterinary, and livestock health. Because the virus was only sporadically detected before 2016 in China, we consider this sudden surge in GETV cases as a re-emerging infectious disease. Therefore, we subsequently performed large-scale GETV PCR-based screening on previously or recently collected samples and found a number of GETV cases starting in 2018. We then sequenced and analyzed the E2 genes of 78 strains (including 16 full genomes) collected from China since 2017 and therefore greatly expanded the existing GETV sequence data. We detailed and demonstrated the advantages of our new approach in assessment of risk unknown disease outbreaks, specifically, we aim at (i) genomic surveillance and analyzing amino acid mutations associated with the ongoing GETV outbreak, (ii) reconstructing the dispersal history of GETV lineages in continental China, (iii) determining which factors were related to the dynamics of viral genetic diversity through time, and (iv) investigating the impact of environmental factors on the dispersal dynamics of GETV

lineages.

Materials and Methods

1. Collection and processing of clinical samples

1.1. Sample collection from dead piglets of unknown etiology

From May 2019 to September 2020, more than 300 piglet deaths of unknown causes occurred at several pig farms in Henan, Guangxi, Hubei and Shandong Province, China. Before the piglets died, they showed clinical symptoms such as diarrhea, wasting, panting, skin rash, and some neurological symptoms. To investigate the cause of dead piglets, we collected swabs, feces and tissue samples of dead piglets from these farms. During the transportation of samples, sufficient cryogenic ice packs were added to the carrying case to maintain a low temperature environment. Collection of all animal samples were approved by the Institutional Animal Care and Use Committee of Nanjing Agricultural University, Nanjing, China (No. SYXK2017-0007).

1.2. Sample processing

Further processing of samples was carried out in a biosafety cabinet. Tissue samples from the lesion area of tissues were cut into small pieces using scissors for surgery and put in 1.5 ML autoclaved tubes. Sterile PBS buffer solution was added after the fecal samples and were divided into equal parts. Swab samples could be directly divided into 200 μl for each sterilized tube and stored. All of the samples stored in the laboratory at -80°C before used.

2. Pathogen identification and retrospective epidemiological survey

2.1. Etiology investigation

According to the clinical symptoms of dead piglets, several routine pathogen detections were implemented on the collected samples, including African swine fever virus (ASFV), porcine reproductive and respiratory syndrome virus (PRRSV), classical swine fever virus (CSFV), pseudorabies virus (PRV), porcine epidemic diarrhea virus

(PEDV), porcine deltacoronavirus (PDCoV), porcine transmissible gastroenteritis virus (TGEV), porcine teschovirus (PTV), porcine kobuvirus (PKV), and porcine circovirus type 2 (PCV2). To further characterize the pathological changes of tissues and organs in dead piglets, we also performed dissections on the piglets.

2.2. Next-generation sequencing

Total RNA was extracted using RNA Clean & Concentrator kit (Zymokit), following the manufacturer's instructions. RNA library was built using TruSeqTM Stranded Total RNA Sample Preparation Kits from Illumina (San Diego, CA) per protocol. Ribo-ZeroTM rRNA Removal Kits from Illumina (San Diego, CA) were used to remove Ribosomal RNA. After fragmented, cDNA synthesis, end repair, A-base addition and Illumina-indexed adaptors ligated, Paired-end (150-bp reads) sequencing of the RNA library was performed on the Novoseq platform (Illumina).

2.3. RNA extraction, pathogen screening and retrospective epidemiological survey of GETV in China

To further investigate the epidemiological situation of GETV in China, after identifying the pathogen based on the results of virus isolation and identification and meta-transcriptome analysis, a retrospective investigation was conducted on samples with similar clinical symptoms collected between 2018 and 2021. Meanwhile, to explore GETV host diversity, we also monitored lab samples from pet dogs and cattle collected over this period. All clinical samples were subsequently screened using primers of Getah virus (GETV) designed according to the available Getah virus sequences of GenBank (www.ncbi.nlm.nih.gov). To obtain GETV genome and *Alphavirus* E2 glycoprotein sequences, the sample RNA was extracted using the OMEGA Viral RNA Isolation Kit (OMEGA, USA), following the manufacturer's instructions strictly. HiScript II 1st Strand cDNA Synthesis Kit (Vazyme, China) was used for cDNA Synthesis. Then, polymerase chain reaction was performed with the GETV detection primers. Samples identified as positive are selected and further conducted amplification

reaction with Phanta® Max Super-Fidelity DNA Polymerase (Vazyme, China) and a set of primers for GETV genome amplification designed based on reference genomes. Subsequently, purified PCR amplification products were sequenced by Sanger dideoxy chain termination method or the NGS method as described in subsection 2.2.

3. Biological characterization of re-emerging GETV strains

Two Getah virus (GETV) isolations were performed from Guangxi province and Henan province. The samples were ground with steel balls under aseptic clean conditions to homogenize. The homogeneous tissues were centrifuged at 16500g for 10 min. The supernatant was filtered through a 0.45μm filter (Millipore, USA), diluted $1:10^5$ with Dulbecco's Modified Eagle Medium (DMEM, Gibco, USA), and then inoculated onto Vero cells cultured in a monolayer. After incubation at 37°C for 1h with 5% CO_2, the inoculum was discarded and maintained in fresh DMEM containing 2% (v/v) fetal bovine serum (FBS, Biological Industries, Israel) and 1% (v/v) penicillin-streptomycin for 48h. Continuous passage like this, when HN isolate passaged to the 5^{th} generation and GX isolate passaged to the 8^{th} generation, a plaque purification assay was performed to purify the virus. Next, virus isolates were confirmed by Reverse Transcription Polymerase Chain Reaction (RT-PCR) and indirect immunofluorescence assay. Then, Virus titration, one-step growth curve determination, immune fluorescence assay (IFA), and mouse infection test of GETV were detail descripted in supplementary materials.

4. Bioinformatic analyses

4.1. Genome assembly

For each library, the Trimmomatic program was used to perform adapter removal and quality trimmed on sequenced reads with the default parameters [38]. Swine and human genomes were collected from NCBI and indexed using BWA and used to map reads with the algorithm “mem”[39]. After removing reads mapped to these genomes, MEGAHIT v1.1.3 was used to assemble the remaining reads [40]. To determine the

potential virus contigs, all of assembled contigs were annotated using diamond based on non-redundant protein database (NR) with 1E-5 E-value cut off. Extracting contigs which were annotated as "Viruses" on kingdom taxonomy lineage information. To estimate the relative abundance of each vertebrate-related virus, unmapped reads were annotated using diamond based on NR databases and we estimated the abundance of each virus as the number of mapped reads per million total reads (RPM) in each library.

4.2. Analysis of genomic sequences

All available GETV genomic sequences and E2 genes from NCBI GenBank database (www.ncbi.nlm.nih.gov) up to December 2, 2021, were collected. Some sequences were removed because (i) they presented duplicated strain names or (ii) corresponded to cell passages of the same original isolate. A total of 159 E2 genes and 59 genomic sequences were used for analysis, including 16 newly obtained genomic sequences and 63 additional newly obtained E2 genes (GenBank accession numbers: MZ736724-MZ736801). E2 is the main protein that mediates virus entry into the host cell during infection, and it is the key point that affects evade immunity and the ability to spread infection [28, 41]. Therefore, during the rapid outbreak and early stage of re-emerging of the virus, we used E2 as molecular markers and preferentially sequenced E2 genes to understand the evolution and transmission of GETV. We performed multiple sequence alignment of the genome and E2 genes using MAFFT (version 7.475) [42] with E-INS-i algorithm, and manually edited the alignment in MEGA (version 7) [43]. Based on the structural and non-structural protein regions of the GETV genome, statistical analysis of the variable amino acid sites was performed.

4.3. Analysis of recombination

We used the program RDP4 4.97 to perform recombination analysis of the genomic sequences and E2 genes [44]. Seven methods LARD [45], 3Seq [46], GENECONV [47], SiScan [48], Chimaera [48], MaxChi [49] and RDP [50] were used to detect recombinant events. We considered that p-values had to be below the 0.05 threshold for

242 at least three of these seven methods to consider the detection of an actual recombinant
243 event [51].
244
245 *4.4 Characterization of selective pressure.*
246 We used the mixed-effects model of evolution (MEME) [52] algorithm in the Hyphy
247 software to estimate non-synonymous to synonymous substitution (dN/dS) ratios. The
248 MEME, single likelihood ancestor counting (SLAC) [53], fast unconstrained Bayesian
249 approximation (FUBAR) [54] and fixed-effects likelihood (FEL) [53] methods were
250 used to estimate the positive selected amino acid sites during evolution. The adaptive
251 branch-site random effects likelihood (aBSREL) method in Hyphy was used to identify
252 specific branches under positive selection during the evolutionary process [55]. When
253 a posterior probability > 0.95 or p-value < 0.05 is met, the site was selected as a
254 potential site; when the criteria of more than 2 method are met, the selected site is
255 considered undergone positive selection. E2 and E1 protein cartoon structures were
256 created with PyMol version 2.1.1 from pdb file 3N40 (Fig. 2A, 2C), 2XFB (2B) or
257 3N41 (2D)[56] which contain the structure of a E1/p62 dimer of CHIKV that shows
258 relatively high amino acid identity (E1: 60%, E2: 54%) and similarity (E1: 77%, E2:
259 68%) to E1 and E2 of GETV.
260
261 *5. Phylogenetic and phylodynamic analyses*
262 *5.1. Phylogenetic and molecular clock analysis of the genomes and E2 gene sequences*
263 A first phylogenetic analysis of both the genome and E2 gene sequences was performed
264 with the maximum likelihood method (ML) implemented in RAxML (version
265 8.2.12)[57] using a general time-reversible nucleotide substitution model with a
266 discretized gamma distribution among sites (GTR + Γ) and 1,000 bootstrap replicates.
267 Temporal signal in our data set was visually assessed using TempEst (1.5.1) [58]. Time-
268 scaled phylogenetic inference was performed using the program BEAST 1.10.4 [59]
269 with high-performance computing library BEAGLE [60]. We assessed the best fitting
270 molecular clock model through marginal likelihood estimation (MLE) using path-

sampling and stepping-stone estimation approaches. With these approaches, we identified the relaxed molecular clock with an underlying lognormal distribution as the best fitting clock model. We specified a GTR+Γ nucleotide substitution model with three partitions for codon positions, an uncorrelated lognormal relaxed molecular clock, and a coalescent-based nonparametric skygrid prior for the tree topologies to model the effective population size over time [16]. Two independent chains with length of 1×10^8 iterations converged to indistinguishable posterior distributions. Convergence and mixing were examined using the program Tracer software 1.7 [61] considering a burn-in of 10% of the total chain length. All parameter estimates yielded effective sample sizes over 200. A maximum clade credibility (MCC) tree summary was generated by TreeAnnotator (1.10.4) and visualized using Figtree 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

5.2. Testing the impact of environmental factors on the viral diversity through time based on the E2 gene

We employed a skygrid- generalized linear model (GLM) coalescent-based model [17] to examine the relationship through time between the viral effective population size and several time-varying covariates. The skygrid-GLM posits a log-linear relationship between the effective population size and a given covariate and enables the inference of an effect size coefficient that quantifies the relationship. Importantly, under the skygrid-GLM model, the effective population size and the effect size coefficient that relates it to a covariate are inferred jointly. This ensures that the effect size coefficient takes the uncertainty in the effective population size reconstruction into account. Finally, in the case of a statistically significant relationship between the effective population size and a covariate, the skygrid-GLM model provides a demographic reconstruction that is based on genetic sequence data as well as the covariate data (in contrast to standard coalescent-based approaches that reconstruct the demographic history exclusively from genetic sequence data). We performed a separate analysis for each of the following five covariates: annual mean temperature, annual precipitation, forest

area, pork production, and per capita meat (pork, beef, mutton) consumption. Annual mean temperature and annual precipitation were retrieved from the WorldClim 2 database (https://worldclim.org/), forest area was taken from the Food and Agriculture Organization of the United Nations (http://www.fao.org/home/en/), and pork production and per capita meat consumption were collected from the National Bureau of Statistics of China (http://www.stats.gov.cn/).

5.3. Phylogeographic analyses based on the E2 gene

We performed both discrete [18] and continuous [19] phylogeographic reconstructions, using the discrete trait and relaxed random walk diffusion models implemented in BEAST 1.10 [59], respectively, and the BEAGLE 3 library [60] to improve computational performance. For both kinds of phylogeographic inference, a distinct reconstruction was performed for each of the two major GETV clades (see below), but within the same BEAST analysis in order to share the estimation of substitution, molecular clock, coalescent, and diffusion model parameters. Specifically, we specified a flexible substitution model with a GTR+Γ parametrization, a relaxed clock model with rates drawn from an underlying lognormal distribution [62], and a flexible nonparametric skygrid coalescent model as the tree topology prior [16].

For the discrete phylogeographic analysis, we used the GLM extension of the discrete diffusion model [21] to jointly infer the dispersal history of lineages among discrete locations as well as the potential contribution of external predictors to the transition rates between pairs of locations. In other words, such a procedure allows investigating the potential impact of external factors on the dispersal frequency of GETV lineages among discrete locations. For each tested predictor, the contribution is estimated by the GLM coefficient, and the associated statistical support is estimated through the computation of a Bayes factor. In the context of this study, we treated the provinces of origin of each sample as discrete locations, and we tested the following predictors using the GLM approach: geographic distance among provinces (great-circle distance between province centroid points; kilometers), pig trade among provinces (Ten

thousand heads/km^2) computed for three different time periods (2017-18, 2017-19, and 2020), the number of pigs slaughtered in the province of origin and the province of destination during two different time periods (2017-18 and 2017-19), the number of pigs raised in the province of origin and the province of destination during two different time periods (2017-18 and 2017-19), and the breeding density of pigs (thousand heads/square kilometer) in the province of origin and the province of destination during two different time periods (2017-18 and 2017-19). In addition, we also included as predictors the numbers of sequences sampled at the province of origin and at the province of destination, which allows to assess the impact of sampling bias on predictor support [21]. For this analysis, a Markov chain Monte Carlo (MCMC) analysis was ran for 10^9 iterations while sampling every 5×10^4 iterations and discarding the first 10% of trees sampled from the posterior distribution as burn-in. Convergence and mixing were examined using the program Tracer 1.7 [61] and all parameter estimates were associated with an estimated sampling size (ESS) greater than 200.

For the continuous phylogeographic analyses, we used a gamma distribution to model the among-branch heterogeneity in diffusion velocity. The MCMC was ran for 2×10^9 iterations while sampling every 10^5 iterations and discarding the first 10% of samples from the posterior distribution as burn-in. Convergence and mixing were again examined using Tracer and all parameter estimates were associated with an ESS greater than 200. MCC trees were summarized using TreeAnnotator 1.10 [59] based on 1,000 trees regularly sampled from the posterior distribution of trees obtained for each of the two major GETV clades considered here. We used R functions available in the package "seraphim" [63, 64] to extract the spatio-temporal information embedded within posterior trees and visualize the dispersal history of GETV lineages and to estimate the weighted lineage dispersal velocity. We further used "seraphim" to perform post hoc analyses of the potential impact of continuous environmental factors on the dispersal location [65] and velocity [66] of viral lineages (Fig. S1): annual mean temperature and annual precipitation retrieved from the WorldClim 2 database (https://worldclim.org/), the pig population density obtained from the Food and Agriculture Organization

database (FAO; http://www.fao.org/livestock-systems/global-distributions/pigs/), the elevation on the study area as estimated by the Shuttle Radar Topography Mission (https://www2.jpl.nasa.gov/srtm), as well as land cover variables (savannas, forests, croplands, urban areas) extracted from land cover data provided by the International Geosphere Biosphere Programme (http://www.igbp.net/).

For investigating the impact of environmental factors on the dispersal location and dispersal velocity of GETV lineages, we applied analytical approaches developed by Dellicour et al. (2019) and (2017), respectively (see Supplementary Information for detailed description of these two approaches).

Results

1. Pathogen identification and retrospective epidemiological survey

All dead pigs in the pig farms of Guangxi, Henan, Hubei and Shandong had suffered from respiratory, digestive and neural symptoms. Therefore, we performed necropsy of the pigs and collected the lung, intestinal and other organs for a pathological section as well as for conventional viral pathogen PCR detection. As shown in Figure 1A, dead pigs exhibit diarrhea, respiratory and neurological symptoms. The autopsy showed emphysema of alveoli, swollen mesenteric lymph nodes, and thinned intestinal walls. None of the common viral pathogens could be identified by PCR, but NGS and bioinformatics analysis identified GETV as the etiological agent (Fig. 1B). Overall, the abundance of GETV was the highest among the four infected farms compared to other viruses, although some GETV positive samples were co-infected with lower concentrations of Picobirnavirus or Kubovirus. In addition, we examined NGS libraries obtained in 2016 before the outbreak from the same GETV-positive swine farms in Guangxi and Shandong provinces, but did not identify any GETV infections. To determine when GETV had re-emerged and started to spread in China, we used Sanger sequencing and NGS to retrospectively analyze laboratory samples collected between 2016 and 2021. In addition to the identification of GETV in laboratory-preserved swine

samples, it is worth noting that we also detected GETV positive nucleic acid and sequenced the GETV E2 gene in mosquitoes, and in cattle and dogs. A total of 78 samples from 16 provinces in China were detected to be positive for GETV. Among them, a total of 16 complete genome (Fig. 1C) sequences were obtained, along with an additional 62 E2 sequences. In addition to the NCBI sequences, we collected case reports of GETV-infections from other labs that did not release any sequences (data not shown). We found that before 2015, there was only one case of swine infection in China, and that was in Henan province in 2012. Since 2015, the number of GETV cases has been increasing year by year, including infections of blue foxes and dogs except livestock, with a wide geographical range of infection (northeast, northwest and the entire south of China). Since 2017, GETV has expanded rapidly in geographical distribution, with mammal cases also appearing in Xinjiang (northwest) and northeast China, but eastern, central and southern China are still the main endemic areas (Fig. S2A). Of note, when the cases are grouped according to seasons, the results show that the number in summer (49 cases, accounted for 47.11%) is significantly higher ($p<0.05$) than in winter (5 cases, 4.81%) (Fig. S2B). This suggests that GETV is more likely to cause disease during the warmer season, when the virus can replicate in mosquito vectors.

2. Biological characterization of re-emerging GETV strains

Two strains of GETV, named HN and GX were isolated by inoculating Vero cells with intestinal abrasive solution after filtration and plaque purification (Fig. 1D). As shown in Figures 1E-H, PK15 or Vero cells inoculated with purified GETV-GX or GETV-HN showed visible cytopathic effect (CPE) in the form of syncytia, rounding and detachment of cells at 48 hpi as compared to the control. A strong signal was observed using anti-E2 antibodies in fluorescence microscopy, indicating that PK15 or Vero cells are effectively infected by GETV-GX or GETV-HN. One-step growth curves demonstrated efficient virus growth in PK15 and Vero cells with virus titers exceeding 10^8 $TCID_{50}$/mL at 48 hpi (Figs. 1I-J).

Of note, GETV can replicate in a variety of animal and primate cell lines, including human cell lines, such as 293T and U251 (Figs. S3A-S3B), which suggests a potential public health risk and human susceptibility to infection. In addition, GETV-GX was also shown to be pathogenic in mouse models. GETV-GX was intracranially inoculated into 3-day-old ICR suckling mice. In the infected group, the weight of the suckling mice ceased to increase after 24 hours. After 48 hours, some suckling mouse began to die with hunched back, tremor, and difficulty in eating. All the suckling mice in the infected group died 80 hours after inoculation (Figs. S3C-S3E).

3. Sequence, mutation and selection analyses

Analysis of all GETV genomes and E2 genes revealed no recombinant signal. More than 50 amino acid substitutions were observed between the recently obtained GETV viruses (data not shown). Here, we focus on 7 amino acid substitutions in E1 and 20 substitutions in E2 relative to a prototype strain, some of which are potential sites under positive selection. Selection pressure analysis revealed that on overall the GETV E2 gene is under purifying selection (date not shown) and only two amino acid sites, E2-86 and E2-323, were found to show evidence for positive selection across all three methods used (FEL, FUBAR, and MEME). In addition, site E2-253 was also found to be subject to positive selection according to the FUBAR analysis (with probability = 0.986), and we found no evidence for positive selection on any individual lineage on the GETV phylogeny. The important mutations and potentially positively selected sites in the ectodomain are highlighted in the crystal structure of the E1/E2 dimer of the closely related Chikungunya virus [56] (Fig. 2A). The four interesting mutations in E2 are also depicted in the trimeric E1/E2 spike, which is shown as a top view (Fig. 2B). Residue 323, which is characterized by a conservative Asp to Glu substitution, is exposed at the surface of the molecule near the membrane. It is thus unlikely to be involved in receptor binding or act as an antibody epitope; its side chain does not form contacts with other amino acids. The substitution His86Tyr is located in the central cavity between E1/E2 dimers, which contains heparan sulfate binding sites in many

445 alphaviruses [67]. The site 207 (Asn207His) is located in a loop at the edge of the spike
446 and exposed at the cell surface (Fig. 2C). This region contains epitopes for cross-
447 reactive neutralizing antibodies which compete with binding to the Mxra8 receptor in
448 other alphaviruses [68, 69]. Residue 253 is located at the base of the viral spike near
449 E3; the side chain of Lys interacts with Tyr 47 of E3 (Fig. 2D). Furthermore, in close
450 vicinity of residue 253 are two other basic amino acids, Arg 250 and Lys251, the latter
451 forms an electrostatic interaction with Asp40 in E3. These amino acids are conserved
452 in other alphaviruses, but Lys is substituted by Arg in GETV variants.
453
454 *4. Phylogenetic, phylodynamic and phylogeographic analyses*
455 By regressing root-to-tip divergences against sampling times, we confirmed the
456 presence of temporal signal for both the whole-genome and E2 gene ML tree using
457 TempEst, with an R^2 of 0.68 and 0.31, respectively. In the absence of clear criteria for
458 genotyping of GETV, we refrain from providing a formal genotype classification.
459 However, based on MCC trees that summarize the time-scaled phylogenetic inference
460 for all genome sequences (Fig. 1C) and E2 sequences (Fig. 2E), we identified 3 lineages
461 with strong posterior support that are responsible for all viruses sampled after 2000,
462 which we refer to as lineage I, II and III. Lineage I has few representatives and contains
463 two mosquito-borne GETV and two swine-borne GETV sequences. Lineage II and III
464 are responsible for the major epidemic strains from pigs in China and GETV from
465 mosquitoes, cattle, blue foxes, horses, lesser panda and canines. To infer the time of
466 emergence of GETV, the time of the most recent common ancestor (TMRCA) of GETV
467 and of each lineage were estimated based on whole genome and E2 sequences. The
468 TMRCA was estimated around 1880 (95% highest posterior density (HPD), [1799,
469 1943]) for the complete genome data set, and around 1904 (95% HPD, [1846, 1947])
470 for the E2 gene. The estimated TMRCA for the three lineages were 1990 (95% HPD
471 [1972, 2000]), 1989 (95% HPD [1976, 2000]) and 1986(95% HPD [1979, 1991]),
472 respectively based on E2. The estimated divergence times for each lineage based on
473 whole genome sequences were similar to the E2 gene estimates. The mean nucleotide

substitution rate (substitutions/site/year) estimated using the whole-genome data set of GETV was 3.19×10^{-4} substitutions/site/year (95% HPD, 2.23×10^{-4} to 4.18×10^{-4}) and using the E2 gene was 6.26×10^{-4} substitutions/site/year (95% HPD, 4.75×10^{-4} to 7.76×10^{-4}). The estimated effective population size of GETV showed that the population diversity of GETV increased year by year since the first outbreak. In addition, the population size of GETV reached its peak around 2018 and maintained a high level until now (Fig. 2F).

Next, we employed a phylodynamic approach to investigate which factors may be associated with the dynamics of GETV genetic diversity through time. The skygrid-GLM analyses clarify the relationship between the viral effective population size and five different covariates. In the case of the per capita meat consumption covariate, we inferred a mean effect size of 0.16 with a 95% Bayesian credibility interval (BCI) of [0.01, 0.30]. Because the 95% BCI excludes zero, we conclude that the relationship between the viral effective population size and per capita meat consumption is statistically significant. On the other hand, all the other skygrid-GLM analyses yielded 95% BCIs that included zero, suggesting that the relationship between the viral effective population size and each of the four remaining covariates falls short of being statistically significant (Fig. S4). In particular, for temperature we inferred a mean effect size of 0.8 with 95% BCI (-0.78, 2.35), for precipitation we inferred a mean effect size of 0.06 with 95% BCI (-0.12, 0.22), for forest area we inferred a mean effect size of 0.06 with 95% BCI (-0.03, 0.14), and for pork production we inferred a mean effect size of 0.0004 with 95% BCI (-0.0009, 0.0016).

The demographic reconstruction resulting from the skygrid-GLM analysis with the per capita meat consumption covariate is shown in Figure 3. The white trajectory represents the mean log viral effective population size and its corresponding 95% BCI region is shown in light blue. This figure also includes per capita meat consumption (shown as red line) as well as a standard skygrid demographic reconstruction that is based only on genetic sequence data (orange mean log effective population size and 95% BCI region

in Figure 3, in contrast to the skygrid-GLM reconstruction that is based on sequence data as well as covariate data. Notably, the mean demographic trajectory inferred by the skygrid-GLM model more closely follows the trajectory of the covariate. Further, the light blue 95% BCI region inferred using the skygrid-GLM is narrower than and almost entirely contained within the orange 95% BCI region inferred using the standard skygrid. In other words, the skygrid-GLM yields a more precise demographic reconstruction that is still compatible with the standard skygrid reconstruction. While meat consumption is certainly not a direct cause of GETV population dynamics, the consumption of meat, such as cattle, sheep, pigs, will greatly impact the frequency and volume of livestock transportation and distribution. This may, for instance, lead to potential contamination of transport vehicles. It is noteworthy that the viral effective population size has a significant association with per capita meat consumption but does not have a significant association with pork production. This suggests that pigs are merely one host for the outbreak and continued epidemic of GETV, and that GETV population dynamics may be due in part to spillover from other species.

The discrete phylogeographic reconstruction coupled with a GLM analysis does not identify support for particular predictors of the dispersal frequency of GETV lineages among Chinese provinces, including live swine trade. Indeed, only the sampling sizes at the province of origin and at the province of destination are associated with Bayes factor values >20, which correspond to strong statistical support [70]. Nevertheless, we found that the Henan province in central China and eastern region in China should be the one of hubs for GETV spread (Fig. S5).

The continuous phylogeographic reconstruction does not allow us to trace the precise origin of the spread of GETV lineages because the uncertainty associated with the location inferred for the root of the tree is relatively pronounced (Fig. 4). However, the reconstructed dispersal history of GETV lineages clearly highlights that some southern and eastern Chinese provinces (Guangxi, Guangdong, Jiangxi, Fujian, Zhejiang) were more recently colonized (>2015; cfr. yellow nodes in Fig. 4). Taking advantage of the continuous phylogeographic reconstruction, we have estimated the weighted dispersal

velocity of GETV lineages: 151.0 km/year (95% HPD = [110.7-203.2]) when considering the entire phylogenetic tree, 139.4 km/year (95% HPD = [99.3-192.3]) when only considering phylogenetic branches occurring before 2015, and 157.8 km/year (95% HPD = [112.2-216.3]) when only considering phylogenetic branches occurring after 2015. While the median value estimated for <2015 is slightly lower than the median value for >2015, their 95% HPD intervals largely overlap. Similarly, we did not identify that more recent (>2015) long-distance lineage dispersal events tended to be associated with relatively higher dispersal velocity (i.e. smaller MCC phylogenetic branch durations for similar geographic distances travelled by those branches; Fig S6). We further tested whether lineage dispersal locations tended to be associated with specific environmental conditions. In practice, we started by computing the E statistic, which measures the mean environmental values at tree node positions. These values were extracted from raster (geo-referenced grids) that summarized the different environmental factors to be tested (Fig. S1). The analyses of the impact of environmental factors on the dispersal location of viral lineages reveal that GETV lineages tend to avoid circulating in areas with higher altitude, and to preferentially circulate within areas associated with relatively higher mean annual temperature and pig population density (Fig. S1).

The analyses of the impact of environmental factors on the dispersal velocity of viral lineages indicate that none of the environmental variables appears to significantly impact the dispersal velocity of GETV lineages: when treated either as conductance or resistance factors and with both path models considered, none of the tested environmental factor is associated with both a positive Q distribution and a Bayes factor support >20. This overall result thus indicates that none of these environmental variables improve the correlation between branch durations and spatial distances (here approximated by the environmental distance computed on a uniform "null" raster), this correlation being already relatively high: R^2 = 0.21 (95% HPD [0.08-0.39]) when spatial distances are computed with the least-cost path model, and R^2 = 0.13 (95% HPD [0.04-0.27]) when spatial distances are computed with the Circuitscape path model. In

other words, our results reveal that, among the environmental factors that we tested, the spatial distance remains the best predictor of the duration associated with GETV lineage dispersal events.

Discussion

Most alphaviruses circulate between specific hematophagous mosquito vectors and susceptible vertebrate hosts, some of which are major public health threats and result in disasters to humans upon spillover [71]. GETV is a member of the *Alphavirus* genus, its pathogenicity for humans is unknown, but there may be a risk of spill-over events to humans [72]. Up to now, epidemiological surveillance studies and available GETV sequences from swine have been rare [26, 73, 74]. In this study, we perform a state-of-the-art genomic surveillance using metagenomic next-generation sequencing coupled with phylodynamic analyses. We find a high abundance of GETV in dead pig samples and identify its link to an outbreak among pig herds in China. We show also that GETV has a broader host range as previously anticipated, which complicates prevention and control because of its diverse reservoir and multiple hosts. We analyze the genetic diversity, dispersal history, and the external factors that may impact the spatial spread of the virus in the early stage of an outbreak/re-emergence in the Chinese pig herd.

We highlight that the current emergence GETV can be divided into three main lineages that primarily evolved and spread in livestock and are geographically widespread. Interestingly, the relatively strong geographical clustering observed in some early mosquito sequences may be related to the limited long-distance travel of mosquitoes or caused by a lack of early sequence samples in livestock, as we found only two lineage I sequence from swine recently. The results of the selection analysis showed that the E2 gene was on overall subject to purifying selection. This is consistent with the widely supported “trade-off” hypothesis for mosquito-borne alphaviruses, i.e. alternate replication in two distinct hosts (vertebrate and invertebrate) limits the evolution of arboviruses, as enhanced fitness in one host may be detrimental to replication in the other host [75]. In addition, the estimated nucleotide substitution rate of GETV is similar to other alphaviruses, such as Ross River virus (RRV) that is most similar to GETV genetically [76]. Of note, we find some evidence for potential adaptive evolution or important amino acid mutations such as H86Y, R253K, N207H in the GETV currently circulating in China. Mapping mutations onto structural models revealed that

two sites might affect binding of GETV to negatively charged heparan sulfate (HS). Different HS-binding sites, basic amino acids, have been identified for equine encephalitis virus (EEV), peripheral sites at the base and axial sites in the central cavity of the viral spike [77]. The selected site His86Tyr in E2 of GETV is also exposed to the central cavity of the viral spike, but the exchange of the weakly basic His by the uncharged Tyr would rather decrease HS-binding. HS-mediated attachment usually increases virus replication in cell culture, but, depending on the virus, either increase or decrease virulence in vivo [67, 78]. The location of the site Lys253Arg corresponds to a peripheral HS-binding site in EEV [77]. Lysine 253 as well as other basic residues in the vicinity interact with amino acids in the E3 subunit. After removal of E3, which detaches from E2 upon virus entry, these basic residues might bind to HS. The K253R substitution, although conservative might directly affect the HS-interaction. Alternatively, it could facilitate or hinder the detachment of E3, which is in other alphaviruses a prerequisite for binding to the cellular receptor and hence for viral infectivity [68, 69]. The other important mutation, Asn207His is located at the surface of E2. Epitopes for broadly neutralizing antibodies, which prevent virus attachment to the Mxra8 receptor are located in the same region [68, 69]. It is unknown whether GETV uses Mxra8 as entry receptor, but other cellular receptors likely bind to the same region in E2 [67]. Importantly, previous examples of epidemic-enhancing mutations in Alphaviruses include CHIKV adaptation to *Ae. albopictus*, and VEEV adaptive mutations that increase replication in horses [79]. Hence, it is possible that GETV may have undergone similar adaptive evolution in Chinese mammals that may lead to public health risk. Therefore, in the wake of the recently sudden outbreak SARS-CoV-2 in human population from un-known animal origin, our results highlight the importance of genome surveillance and early warning of emerging infectious diseases in animals. Moreover, our study has enabled a more robust analysis of GETV evolutionary history and revealed a more extensive genetic diversity compared to previous analyses [28]. We estimate that the overall genetic diversity of GETV has increased through time since the first report, which increased the possibility of a large-scale GETV outbreak [80].

GETV has a wide geographical distribution in China, especially in the southern region as well as in areas used for livestock (Fig. S1), which might be related to the distribution of its mosquito vector. Therefore, we examined the trajectory of GETV effective population size over time and compared it to a number of different factors that have been hypothesized to be related to GETV population dynamics. Of note, we found a statistically significant positive association between the viral effective population size and per capita livestock meat consumption. This result suggests that it may not simply be the breeding and trade of swine themselves that caused the GETV outbreak, but rather that GETV may have been prevalent in other livestock for some time and may have been partly responsible for causing the outbreak in pigs via mosquitos. This hypothesis needs to be explored in more detail in future work.

Finally, by testing the association between environmental factors and the locations of lineage dispersal, we demonstrate that, overall, GETV lineages have preferentially circulated in specific environmental conditions (higher temperature) and in regions with higher swine population density. In this respect, it is important to note that the highest mosquito density in China occurs near livestock farms [81] and that GETV incidence in mammals is significantly higher in summer than in winter. On the other hand, differential sampling efforts may bias association estimates with environmental factors, so we take these finds as more suggestive than conclusive. Nonetheless, our results provide insight into the evolution and diffusion of GETV that may help to prevent and control GETV infections in livestock. We recommend increased sample collection from and surveillance of a wider range of species and geographic regions, as uncovering the transmission routes and major sources of GETV in animals will help prevent future outbreaks of GETV disease among livestock and emergences in humans.

Our findings, as alluded to above, should be interpreted in the light of particular limitations. Uneven sampling may affect our results, and although our pig sampling covers well the sites we surveille, we may lack a large number of samples from other sites as well as from other livestock and wildlife animals. Furthermore, our

phylogeographic analyses are strongly influenced by the sampling effort, and therefore remain somehow more descriptive of the environmental conditions associated with the dispersal locations of inferred viral lineages (Dellicour, et al. 2019).

Our research is the first to integrate transcriptomics, genomics, genomic epidemiology and landscape epidemiology, revealing that the unexplained pig herd deaths were caused by multiple different lineages of re-emerging GETV in China. Our ability to predict future pandemics will require intensified viral surveillance and an understanding of how economic forces and livestock trade policies affect changes in animal movements and production practices that drive viral emergence. We also demonstrate that usage of modern technological platforms, such as NGS and phylogenetic analysis, allows to identify virus outbreaks more rapidly than traditional methods, such as PCR and virus isolation. Furthermore, our study suggests for the first time that GETV could have the potential to emerge in human populations, especially in areas with high temperature and high livestock production in China, due to the accumulation of mutations and its high genetic diversity and wide host range.

Acknowledge

Shuo Su released preprints on behalf of all authors.

Reference

1. Beigel JH, Farrar J, Han AM, Hayden FG, Hyer R, de Jong MD, et al. Avian influenza A (H5N1) infection in humans. N Engl J Med. 2005;353(13):1374-85. Epub 2005/09/30. doi: 10.1056/NEJMra052211. PubMed PMID: 16192482.

2. Li W, Shi Z, Yu M, Ren W, Smith C, Epstein JH, et al. Bats are natural reservoirs of SARS-like coronaviruses. Science. 2005;310(5748):676-9. Epub 2005/10/01. doi: 10.1126/science.1118391. PubMed PMID: 16195424.

3. Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol. 2019;17(3):181-92. Epub 2018/12/12. doi: 10.1038/s41579-018-0118-9. PubMed PMID: 30531947; PubMed Central PMCID: PMCPMC7097006.

4. Wang LF, Anderson DE. Viruses in bats and potential spillover to animals and humans. Curr Opin Virol. 2019;34:79-89. Epub 2019/01/22. doi: 10.1016/j.coviro.2018.12.007. PubMed PMID: 30665189; PubMed Central PMCID: PMCPMC7102861.

5. Sun J, He WT, Wang L, Lai A, Ji X, Zhai X, et al. COVID-19: Epidemiology, Evolution, and Cross-Disciplinary Perspectives. Trends Mol Med. 2020;26(5):483-95. Epub 2020/05/04. doi:

10.1016/j.molmed.2020.02.008. PubMed PMID: 32359479; PubMed Central PMCID: PMCPMC7118693.

6. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;579(7798):270-3. Epub 2020/02/06. doi: 10.1038/s41586-020-2012-7. PubMed PMID: 32015507; PubMed Central PMCID: PMCPMC7095418.

7. He W, Auclert LZ, Zhai X, Wong G, Zhang C, Zhu H, et al. Interspecies Transmission, Genetic Diversity, and Evolutionary Dynamics of Pseudorabies Virus. J Infect Dis. 2019;219(11):1705-15. Epub 2018/12/28. doi: 10.1093/infdis/jiy731. PubMed PMID: 30590733.

8. He WT, Ji X, He W, Dellicour S, Wang S, Li G, et al. Genomic Epidemiology, Evolution, and Transmission Dynamics of Porcine Deltacoronavirus. Mol Biol Evol. 2020;37(9):2641-54. Epub 2020/05/15. doi: 10.1093/molbev/msaa117. PubMed PMID: 32407507; PubMed Central PMCID: PMCPMC7454817.

9. Ang BSP, Lim TCC, Wang L. Nipah Virus Infection. J Clin Microbiol. 2018;56(6). Epub 2018/04/13. doi: 10.1128/JCM.01875-17. PubMed PMID: 29643201; PubMed Central PMCID: PMCPMC5971524.

10. Novel Swine-Origin Influenza AVIT, Dawood FS, Jain S, Finelli L, Shaw MW, Lindstrom S, et al. Emergence of a novel swine-origin influenza A (H1N1) virus in humans. N Engl J Med. 2009;360(25):2605-15. Epub 2009/05/09. doi: 10.1056/NEJMoa0903810. PubMed PMID: 19423869.

11. Mena I, Nelson MI, Quezada-Monroy F, Dutta J, Cortes-Fernandez R, Lara-Puente JH, et al. Origins of the 2009 H1N1 influenza pandemic in swine in Mexico. Elife. 2016;5. Epub 2016/06/29. doi: 10.7554/eLife.16777. PubMed PMID: 27350259; PubMed Central PMCID: PMCPMC4957980.

12. Liu Q, Wang X, Xie C, Ding S, Yang H, Guo S, et al. A novel human acute encephalitis caused by pseudorabies virus variant strain. Clin Infect Dis. 2020. Epub 2020/07/16. doi: 10.1093/cid/ciaa987. PubMed PMID: 32667972.

13. Bragg L, Tyson GW. Metagenomics using next-generation sequencing. Methods Mol Biol. 2014;1096:183-201. Epub 2014/02/12. doi: 10.1007/978-1-62703-712-9_15. PubMed PMID: 24515370.

14. Miao Q, Ma Y, Wang Q, Pan J, Zhang Y, Jin W, et al. Microbiological Diagnostic Performance of Metagenomic Next-generation Sequencing When Applied to Clinical Practice. Clin Infect Dis. 2018;67(suppl_2):S231-S40. Epub 2018/11/14. doi: 10.1093/cid/ciy693. PubMed PMID: 30423048.

15. Kalantar KL, Carvalho T, de Bourcy CFA, Dimitrov B, Dingle G, Egger R, et al. IDseq-An open source cloud-based pipeline and analysis service for metagenomic pathogen detection and monitoring. GigaScience. 2020;9(10). Epub 2020/10/16. doi: 10.1093/gigascience/giaa111. PubMed PMID: 33057676; PubMed Central PMCID: PMCPMC7566497.

16. Gill MS, Lemey P, Faria NR, Rambaut A, Shapiro B, Suchard MA. Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. Mol Biol Evol. 2013;30(3):713-24. Epub 2012/11/28. doi: 10.1093/molbev/mss265. PubMed PMID: 23180580; PubMed Central PMCID: PMCPMC3563973.

17. Gill MS, Lemey P, Bennett SN, Biek R, Suchard MA. Understanding Past Population Dynamics: Bayesian Coalescent-Based Modeling with Covariates. Syst Biol. 2016;65(6):1041-56. Epub 2016/07/03. doi: 10.1093/sysbio/syw050. PubMed PMID: 27368344; PubMed Central PMCID: PMCPMC5066065.

728 18. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots.
729 PLoS Comput Biol. 2009;5(9):e1000520. Epub 2009/09/26. doi: 10.1371/journal.pcbi.1000520.
730 PubMed PMID: 19779555; PubMed Central PMCID: PMCPMC2740835.
731 19. Lemey P, Rambaut A, Welch JJ, Suchard MA. Phylogeography takes a relaxed random walk in
732 continuous space and time. Mol Biol Evol. 2010;27(8):1877-85. Epub 2010/03/06. doi:
733 10.1093/molbev/msq067. PubMed PMID: 20203288; PubMed Central PMCID: PMCPMC2915639.
734 20. Pybus OG, Suchard MA, Lemey P, Bernardin FJ, Rambaut A, Crawford FW, et al. Unifying the
735 spatial epidemiology and molecular evolution of emerging epidemics. Proc Natl Acad Sci U S A.
736 2012;109(37):15066-71. Epub 2012/08/29. doi: 10.1073/pnas.1206598109. PubMed PMID:
737 22927414; PubMed Central PMCID: PMCPMC3443149.
738 21. Lemey P, Rambaut A, Bedford T, Faria N, Bielejec F, Baele G, et al. Unifying viral genetics and
739 human transportation data to predict the global transmission dynamics of human influenza H3N2.
740 PLoS Pathog. 2014;10(2):e1003932. Epub 2014/03/04. doi: 10.1371/journal.ppat.1003932.
741 PubMed PMID: 24586153; PubMed Central PMCID: PMCPMC3930559.
742 22. Jacquot M, Nomikou K, Palmarini M, Mertens P, Biek R. Bluetongue virus spread in Europe is
743 a consequence of climatic, landscape and vertebrate host factors as revealed by phylogeographic
744 inference. Proc Biol Sci. 2017;284(1864). Epub 2017/10/13. doi: 10.1098/rspb.2017.0919. PubMed
745 PMID: 29021180; PubMed Central PMCID: PMCPMC5647287.
746 23. Brunker K, Lemey P, Marston DA, Fooks AR, Lugelo A, Ngeleja C, et al. Landscape attributes
747 governing local transmission of an endemic zoonosis: Rabies virus in domestic dogs. Mol Ecol.
748 2018;27(3):773-88. Epub 2017/12/24. doi: 10.1111/mec.14470. PubMed PMID: 29274171;
749 PubMed Central PMCID: PMCPMC5900915.
750 24. Rawle DJ, Nguyen W, Dumenil T, Parry R, Warrilow D, Tang B, et al. Sequencing of Historical
751 Isolates, K-mer Mining and High Serological Cross-Reactivity with Ross River Virus Argue against
752 the Presence of Getah Virus in Australia. Pathogens. 2020;9(10). Epub 2020/10/22. doi:
753 10.3390/pathogens9100848. PubMed PMID: 33081269; PubMed Central PMCID:
754 PMCPMC7650646.
755 25. Kumanomido T, Wada R, Kanemaru T, Kamada M, Hirasawa K, Akiyama Y. Clinical and
756 virological observations on swine experimentally infected with Getah virus. Vet Microbiol.
757 1988;16(3):295-301. Epub 1988/03/01. doi: 10.1016/0378-1135(88)90033-8. PubMed PMID:
758 2836997.
759 26. Xing C, Jiang J, Lu Z, Mi S, He B, Tu C, et al. Isolation and characterization of Getah virus from
760 pigs in Guangdong province of China. Transboundary and emerging diseases. 2020. Epub
761 2020/04/12. doi: 10.1111/tbed.13567. PubMed PMID: 32277601.
762 27. Zhai YG, Wang HY, Sun XH, Fu SH, Wang HQ, Attoui H, et al. Complete sequence
763 characterization of isolates of Getah virus (genus Alphavirus, family Togaviridae) from China. J Gen
764 Virol. 2008;89(Pt 6):1446-56. Epub 2008/05/14. doi: 10.1099/vir.0.83607-0. PubMed PMID:
765 18474561.
766 28. Li YY, Liu H, Fu SH, Li XL, Guo XF, Li MH, et al. From discovery to spread: The evolution and
767 phylogeny of Getah virus. Infect Genet Evol. 2017;55:48-55. Epub 2017/08/23. doi:
768 10.1016/j.meegid.2017.08.016. PubMed PMID: 28827175.
769 29. Liu H, Zhang X, Li LX, Shi N, Sun XT, Liu Q, et al. First isolation and characterization of Getah
770 virus from cattle in northeastern China. BMC veterinary research. 2019;15(1):320. Epub 2019/09/07.
771 doi: 10.1186/s12917-019-2061-z. PubMed PMID: 31488162; PubMed Central PMCID:

772 PMCPMC6729113.
773 30. Lu G, Ou J, Ji J, Ren Z, Hu X, Wang C, et al. Emergence of Getah Virus Infection in Horse With
774 Fever in China, 2018. Frontiers in microbiology. 2019;10:1416. Epub 2019/07/10. doi:
775 10.3389/fmicb.2019.01416. PubMed PMID: 31281304; PubMed Central PMCID: PMCPMC6596439.
776 31. Li L, Guo X, Zhao Q, Tong Y, Fan H, Sun Q, et al. Investigation on Mosquito-Borne Viruses at
777 Lancang River and Nu River Watersheds in Southwestern China. Vector borne and zoonotic
778 diseases (Larchmont, NY). 2017;17(12):804-12. Epub 2017/10/31. doi: 10.1089/vbz.2017.2164.
779 PubMed PMID: 29083983.
780 32. Shi N, Li LX, Lu RG, Yan XJ, Liu H. Highly Pathogenic Swine Getah Virus in Blue Foxes, Eastern
781 China, 2017. Emerging infectious diseases. 2019;25(6):1252-4. Epub 2019/05/21. doi:
782 10.3201/eid2506.181983. PubMed PMID: 31107236; PubMed Central PMCID: PMCPMC6537705.
783 33. Li Y, Fu S, Guo X, Li X, Li M, Wang L, et al. Serological Survey of Getah Virus in Domestic
784 Animals in Yunnan Province, China. Vector borne and zoonotic diseases (Larchmont, NY).
785 2019;19(1):59-61. Epub 2018/06/30. doi: 10.1089/vbz.2018.2273. PubMed PMID: 29957135.
786 34. Doherty RL, Gorman BM, Whitehead RH, Carley JG. Studies of arthropod-borne virus
787 infections in Queensland. V. Survey of antibodies to group A arboviruses in man and other animals.
788 The Australian journal of experimental biology and medical science. 1966;44(4):365-77. Epub
789 1966/08/01. doi: 10.1038/icb.1966.35. PubMed PMID: 6007741.
790 35. Taylor KG, Paessler S. Pathogenesis of Venezuelan equine encephalitis. Vet Microbiol.
791 2013;167(1-2):145-50. Epub 2013/08/24. doi: 10.1016/j.vetmic.2013.07.012. PubMed PMID:
792 23968890.
793 36. Lu G, Chen R, Shao R, Dong N, Liu W, Li S. Getah virus: An increasing threat in China. J Infect.
794 2020;80(3):350-71. Epub 2019/12/04. doi: 10.1016/j.jinf.2019.11.016. PubMed PMID: 31790706.
795 37. Suhrbier A. Rheumatic manifestations of chikungunya: emerging concepts and interventions.
796 Nat Rev Rheumatol. 2019;15(10):597-611. Epub 2019/09/05. doi: 10.1038/s41584-019-0276-9.
797 PubMed PMID: 31481759.
798 38. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data.
799 Bioinformatics. 2014;30(15):2114-20. Epub 2014/04/04. doi: 10.1093/bioinformatics/btu170.
800 PubMed PMID: 24695404; PubMed Central PMCID: PMCPMC4103590.
801 39. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:
802 Genomics. 2013.
803 40. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for
804 large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics.
805 2015;31(10):1674-6. Epub 2015/01/23. doi: 10.1093/bioinformatics/btv033. PubMed PMID:
806 25609793.
807 41. Tsetsarkin KA, Weaver SC. Sequential adaptive mutations enhance efficient vector switching
808 by Chikungunya virus and its epidemic emergence. PLoS Pathog. 2011;7(12):e1002412. Epub
809 2011/12/17. doi: 10.1371/journal.ppat.1002412. PubMed PMID: 22174678; PubMed Central
810 PMCID: PMCPMC3234230.
811 42. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements
812 in performance and usability. Molecular biology and evolution. 2013;30(4):772-80. Epub
813 2013/01/19. doi: 10.1093/molbev/mst010. PubMed PMID: 23329690; PubMed Central PMCID:
814 PMCPMC3603318.
815 43. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0

816 for Bigger Datasets. Mol Biol Evol. 2016;33(7):1870-4. Epub 2016/03/24. doi:
817 10.1093/molbev/msw054. PubMed PMID: 27004904; PubMed Central PMCID: PMCPMC8210823.
818 44. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: Detection and analysis of
819 recombination patterns in virus genomes. Virus evolution. 2015;1(1):vev003. Epub 2015/05/26. doi:
820 10.1093/ve/vev003. PubMed PMID: 27774277; PubMed Central PMCID: PMCPMC5014473.
821 45. Holmes EC, Worobey M, Rambaut A. Phylogenetic evidence for recombination in dengue
822 virus. Molecular biology and evolution. 1999;16(3):405-9. Epub 1999/05/20. doi:
823 10.1093/oxfordjournals.molbev.a026121. PubMed PMID: 10331266.
824 46. Boni MF, Posada D, Feldman MW. An exact nonparametric method for inferring mosaic
825 structure in sequence triplets. Genetics. 2007;176(2):1035-47. Epub 2007/04/06. doi:
826 10.1534/genetics.106.068874. PubMed PMID: 17409078; PubMed Central PMCID:
827 PMCPMC1894573.
828 47. Padidam M, Sawyer S, Fauquet CM. Possible emergence of new geminiviruses by frequent
829 recombination. Virology. 1999;265(2):218-25. Epub 1999/12/22. doi: 10.1006/viro.1999.0056.
830 PubMed PMID: 10600594.
831 48. Gibbs MJ, Armstrong JS, Gibbs AJ. Sister-scanning: a Monte Carlo procedure for assessing
832 signals in recombinant sequences. Bioinformatics (Oxford, England). 2000;16(7):573-82. Epub
833 2000/10/20. doi: 10.1093/bioinformatics/16.7.573. PubMed PMID: 11038328.
834 49. Smith JM. Analyzing the mosaic structure of genes. Journal of molecular evolution.
835 1992;34(2):126-9. Epub 1992/02/01. doi: 10.1007/bf00182389. PubMed PMID: 1556748.
836 50. Martin D, Rybicki E. RDP: detection of recombination amongst aligned sequences.
837 Bioinformatics (Oxford, England). 2000;16(6):562-3. Epub 2000/09/12. doi:
838 10.1093/bioinformatics/16.6.562. PubMed PMID: 10980155.
839 51. Sabir JS, Lam TT, Ahmed MM, Li L, Shen Y, Abo-Aba SE, et al. Co-circulation of three camel
840 coronavirus species and recombination of MERS-CoVs in Saudi Arabia. Science (New York, NY).
841 2016;351(6268):81-4. Epub 2015/12/19. doi: 10.1126/science.aac8608. PubMed PMID: 26678874.
842 52. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting
843 individual sites subject to episodic diversifying selection. PLoS Genet. 2012;8(7):e1002764. Epub
844 2012/07/19. doi: 10.1371/journal.pgen.1002764. PubMed PMID: 22807683; PubMed Central
845 PMCID: PMCPMC3395634.
846 53. Kosakovsky Pond SL, Frost SD. Not so different after all: a comparison of methods for
847 detecting amino acid sites under selection. Mol Biol Evol. 2005;22(5):1208-22. Epub 2005/02/11.
848 doi: 10.1093/molbev/msi105. PubMed PMID: 15703242.
849 54. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, et al. FUBAR: a
850 fast, unconstrained bayesian approximation for inferring selection. Mol Biol Evol. 2013;30(5):1196-
851 205. Epub 2013/02/20. doi: 10.1093/molbev/mst030. PubMed PMID: 23420840; PubMed Central
852 PMCID: PMCPMC3670733.
853 55. Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL. Less is more:
854 an adaptive branch-site random effects model for efficient detection of episodic diversifying
855 selection. Mol Biol Evol. 2015;32(5):1342-53. Epub 2015/02/24. doi: 10.1093/molbev/msv022.
856 PubMed PMID: 25697341; PubMed Central PMCID: PMCPMC4408413.
857 56. Voss JE, Vaney MC, Duquerroy S, Vonrhein C, Girard-Blanc C, Crublet E, et al. Glycoprotein
858 organization of Chikungunya virus particles revealed by X-ray crystallography. Nature.
859 2010;468(7324):709-12. Epub 2010/12/03. doi: 10.1038/nature09555. PubMed PMID: 21124458.

860 57. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
861 phylogenies. Bioinformatics. 2014;30(9):1312-3. Epub 2014/01/24. doi:
862 10.1093/bioinformatics/btu033. PubMed PMID: 24451623; PubMed Central PMCID:
863 PMCPMC3998144.
864 58. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of
865 heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. 2016;2(1):vew007.
866 Epub 2016/10/25. doi: 10.1093/ve/vew007. PubMed PMID: 27774300; PubMed Central PMCID:
867 PMCPMC4989882.
868 59. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic
869 and phylodynamic data integration using BEAST 1.10. Virus Evol. 2018;4(1):vey016. Epub
870 2018/06/27. doi: 10.1093/ve/vey016. PubMed PMID: 29942656; PubMed Central PMCID:
871 PMCPMC6007674.
872 60. Ayres DL, Cummings MP, Baele G, Darling AE, Lewis PO, Swofford DL, et al. BEAGLE 3:
873 Improved Performance, Scaling, and Usability for a High-Performance Computing Library for
874 Statistical Phylogenetics. Syst Biol. 2019;68(6):1052-61. Epub 2019/04/30. doi:
875 10.1093/sysbio/syz020. PubMed PMID: 31034053; PubMed Central PMCID: PMCPMC6802572.
876 61. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior Summarization in Bayesian
877 Phylogenetics Using Tracer 1.7. Syst Biol. 2018;67(5):901-4. Epub 2018/05/03. doi:
878 10.1093/sysbio/syy032. PubMed PMID: 29718447; PubMed Central PMCID: PMCPMC6101584.
879 62. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with
880 confidence. PLoS Biol. 2006;4(5):e88. Epub 2006/05/11. doi: 10.1371/journal.pbio.0040088.
881 PubMed PMID: 16683862; PubMed Central PMCID: PMCPMC1395354.
882 63. Dellicour S, Rose R, Pybus OG. Explaining the geographic spread of emerging epidemics: a
883 framework for comparing viral phylogenies and environmental landscape data. BMC
884 Bioinformatics. 2016;17:82. Epub 2016/02/13. doi: 10.1186/s12859-016-0924-x. PubMed PMID:
885 26864798; PubMed Central PMCID: PMCPMC4750353.
886 64. Dellicour S, Rose R, Faria NR, Lemey P, Pybus OG. SERAPHIM: studying environmental rasters
887 and phylogenetically informed movements. Bioinformatics. 2016;32(20):3204-6. Epub 2016/06/24.
888 doi: 10.1093/bioinformatics/btw384. PubMed PMID: 27334476.
889 65. Dellicour S, Troupin C, Jahanbakhsh F, Salama A, Massoudi S, Moghaddam MK, et al. Using
890 phylogeographic approaches to analyse the dispersal history, velocity and direction of viral
891 lineages — Application to rabies virus spread in Iran. Molecular Ecology. 2019;28(18):4335-50. doi:
892 https://doi.org/10.1111/mec.15222.
893 66. Dellicour S, Rose R, Faria NR, Vieira LFP, Bourhy H, Gilbert M, et al. Using Viral Gene Sequences
894 to Compare and Explain the Heterogeneous Spatial Dynamics of Virus Epidemics. Molecular
895 Biology and Evolution. 2017;34(10):2563-71. doi: 10.1093/molbev/msx176.
896 67. Holmes AC, Basore K, Fremont DH, Diamond MS. A molecular understanding of alphavirus
897 entry. PLoS Pathog. 2020;16(10):e1008876. Epub 2020/10/23. doi: 10.1371/journal.ppat.1008876.
898 PubMed PMID: 33091085; PubMed Central PMCID: PMCPMC7580943 the Scientific Advisory
899 Board of Moderna. The Diamond laboratory has received unrelated funding under sponsored
900 research agreements from Vir Biotechnology, Moderna, and Emergent BioSolutions.
901 68. Fox JM, Long F, Edeling MA, Lin H, van Duijl-Richter MKS, Fong RH, et al. Broadly Neutralizing
902 Alphavirus Antibodies Bind an Epitope on E2 and Inhibit Entry and Egress. Cell. 2015;163(5):1095-
903 107. Epub 2015/11/11. doi: 10.1016/j.cell.2015.10.050. PubMed PMID: 26553503; PubMed Central

PMCID: PMCPMC4659373.

69. Powell LA, Miller A, Fox JM, Kose N, Klose T, Kim AS, et al. Human mAbs Broadly Protect against Arthritogenic Alphaviruses by Recognizing Conserved Elements of the Mxra8 Receptor-Binding Site. Cell Host Microbe. 2020;28(5):699-711 e7. Epub 2020/08/14. doi: 10.1016/j.chom.2020.07.008. PubMed PMID: 32783883; PubMed Central PMCID: PMCPMC7666055.

70. Kass RE, Raftery AE. Bayes Factors. Journal of the American Statistical Association. 1995;90(430):773-95. doi: 10.1080/01621459.1995.10476572.

71. Azar SR, Campos RK, Bergren NA, Camargos VN, Rossi SL. Epidemic Alphaviruses: Ecology, Emergence and Outbreaks. Microorganisms. 2020;8(8):1167. doi: 10.3390/microorganisms8081167. PubMed PMID: 32752150.

72. Guth S, Hanley KA, Althouse BM, Boots M. Ecological processes underlying the emergence of novel enzootic cycles: Arboviruses in the neotropics as a case study. PLoS neglected tropical diseases. 2020;14(8):e0008338-e. doi: 10.1371/journal.pntd.0008338. PubMed PMID: 32790670.

73. Ren T, Mo Q, Wang Y, Wang H, Nong Z, Wang J, et al. Emergence and Phylogenetic Analysis of a Getah Virus Isolated in Southern China. Front Vet Sci. 2020;7:552517. Epub 2020/12/22. doi: 10.3389/fvets.2020.552517. PubMed PMID: 33344520; PubMed Central PMCID: PMCPMC7744783.

74. Rattanatumhi K, Prasertsincharoen N, Naimon N, Kuwata R, Shimoda H, Ishijima K, et al. A serological survey and characterization of Getah virus in domestic pigs in Thailand, 2017-2018. Transboundary and emerging diseases. 2021. Epub 2021/02/23. doi: 10.1111/tbed.14042. PubMed PMID: 33617130.

75. Althouse BM, Hanley KA. The tortoise or the hare? Impacts of within-host dynamics on transmission success of arthropod-borne viruses. Philos Trans R Soc Lond B Biol Sci. 2015;370(1675). Epub 2015/07/08. doi: 10.1098/rstb.2014.0299. PubMed PMID: 26150665; PubMed Central PMCID: PMCPMC4528497.

76. Michie A, Dhanasekaran V, Lindsay MDA, Neville PJ, Nicholson J, Jardine A, et al. Genome-Scale Phylogeny and Evolutionary Analysis of Ross River Virus Reveals Periodic Sweeps of Lineage Dominance in Western Australia, 1977-2014. J Virol. 2020;94(2). Epub 2019/11/02. doi: 10.1128/JVI.01234-19. PubMed PMID: 31666378; PubMed Central PMCID: PMCPMC6955267.

77. Chen CL, Hasan SS, Klose T, Sun Y, Buda G, Sun C, et al. Cryo-EM structure of eastern equine encephalitis virus in complex with heparan sulfate analogues. Proc Natl Acad Sci U S A. 2020;117(16):8890-9. Epub 2020/04/05. doi: 10.1073/pnas.1910670117. PubMed PMID: 32245806; PubMed Central PMCID: PMCPMC7183182.

78. Button JM, Qazi SA, Wang JC, Mukhopadhyay S. Revisiting an old friend: new findings in alphavirus structure and assembly. Curr Opin Virol. 2020;45:25-33. Epub 2020/07/20. doi: 10.1016/j.coviro.2020.06.005. PubMed PMID: 32683295; PubMed Central PMCID: PMCPMC7746636.

79. Brault AC, Powers AM, Holmes EC, Woelk CH, Weaver SC. Positively charged amino acid substitutions in the e2 envelope glycoprotein are associated with the emergence of venezuelan equine encephalitis virus. J Virol. 2002;76(4):1718-30. Epub 2002/01/19. doi: 10.1128/jvi.76.4.1718-1730.2002. PubMed PMID: 11799167; PubMed Central PMCID: PMCPMC135911.

80. Atoni E, Zhao L, Hu C, Ren N, Wang X, Liang M, et al. A dataset of distribution and diversity of mosquito-associated viruses and their mosquito vectors in China. Sci Data. 2020;7(1):342. Epub

948 2020/10/15. doi: 10.1038/s41597-020-00687-9. PubMed PMID: 33051449; PubMed Central
949 PMCID: PMCPMC7555486.
950 81. Wu H, Lu L, Meng F, Guo Y, QY L. Reports on national surveillance of mosquitoes in China,
951 2006-2015. Chin J Vector Biol & Control. 2017;28(005):409-15.

952

953 **Figure Legends**

954 **Figure 1**. Isolation and characterization of Getah virus (GETV). (A) GETV-infected
955 piglets showing clinical features. From left to right: cyanosis, diarrhea, thinning of the
956 intestinal wall and lymphadenopathy. (B) The abundance of various viruses at the genus
957 level in GETV-positive and -negative farms. The relative abundance of each virus in
958 each library was estimated and normalized by the number of mapped reads per million
959 total reads (RPM). To remove contaminations, we only show RPM above 1. Guangxi-
960 2019, Henan-2020, Shandong-2019 and Hubei-2019: GETV positive farms. Guangxi-
961 2016 and Shandong- 2016 corresponded to the two farms that were GETV-positive in
962 2019. (C) Maximum clade credibility tree of GETV based on whole genome sequences.
963 Red squares represent new sequences obtained in this study. The sampling location and
964 the host are color-coded. (D) GETV was successfully isolated and verified by agarose
965 gel electrophoresis. (E-F) Vero were infected with GETV-GX or GETV-HN
966 (MOI=0.001) and PK15 cells were infected with GETV-GX or GETV-HN (MOI=0.1).
967 Cytopathic changes was observed at 12, 24, 36, 48, 60 and 72 hpi. (G-H)
968 Immunofluorescence of GETV-E2 (green) detected in infected Vero or PK15 cells,
969 respectively, Nuclei are stained blue with DAPI. All fluorescent images were taken at
970 20×magnification. (I-J) Growth of GETV-GX or GETV-HN in Vero (I) and PK15 (J)
971 cultures. Viral titers were determined for samples (only medium) between 12 and 72
972 hpi in Vero cells. Data are expressed as mean ±S.D. of viral titers (lg10 $TCID_{50}$ per
973 0.1ml) derived from three infected cell cultures.

974

975 **Figure 2**. Location of amino acid substitutions and selected sites in E2 of GETV
976 variants. (A): Structure of a heterodimer containing the E1 (green cartoon) and E2 (blue)
977 subunit. The small E3 subunit (magenta cartoon) is still associated with E2. Amino acid
978 exchanges are highlighted as red spheres. The horizontal line symbolizes the viral

membrane, in which both proteins are anchored by a transmembrane region. FL = fusion loop in E1. (B) Top view of a hexameric spike composed of three E1 (green cartoon) and three E2 (blue) subunits. Positively selected and other interesting sites in E2 are highlighted as red spheres. (C) Detail of the E2 structure in a semitransparent surface projection showing location of residue 207 as red stick. Epitopes for antibodies that prevent binding of alphaviruses to the Mxra8 receptor are shown as orange sticks [69] and for other broadly neutralizing antibodies as wheat sticks [68]. (D) Detail of the interface of E2 (blue) with E3 (green) showing the location of the selected site 253 as a stick. K253 interacts with Tyr 47 in the E3 subunit. Shown as sticks are also two other basic amino acids in the vicinity, one of them (R250) forms an ionic interaction with D40 in E3. After removal of E3 during virus entry, the three basic amino acids might from a heparan sulphate (HS) binding site. The conservative exchange K253R might affect HS-binding or removal of E3. The figures were created with PyMol from pdb-files 3N40 (a,c,d) or 2XFB (b) (E) Maximum clade credibility tree (MCC) of GETV E2 gene obtained from time-scaled phylogenetic inference. F) the effective population size over time of GETV E2 gene with an uncorrelated lognormal relaxed molecular clock, and a coalescent-based nonparametric skygrid prior for the tree topologies.

Figure 3. Skygrid demographic reconstructions. The dark orange line and shaded light orange region represent the mean log viral effective population size and its 95% Bayesian credibility interval (BCI) region, respectively, inferred using a standard skygrid analysis of sequence data. The white line and shaded light blue region represent the mean log viral effective population size and its 95% BCI region, respectively, inferred using a skygrid-GLM analysis of sequence data and per capita meat (pork, beef, mutton) consumption covariate data. The per capita meat consumption is shown in red. The skygrid-GLM analysis yields an effect size coefficient with mean 0.16 and 95% BCI (0.01, 0.30), indicating a statistically significant association between the viral effective population size and per capita meat consumption.

Figure 4. Dispersal history of GETV lineages in China: maximum clade credibility (MCC) tree and 80% highest posterior density (HPD) regions reflecting the uncertainty related to the phylogeographic inference and based on 1,000 trees subsampled from the posterior distribution. MCC tree nodes are colored according to their time of occurrence, and 80% HPD regions were computed for successive time layers and then superimposed using the same color scale reflecting time. In addition to the overall continuous phylogeographic reconstruction, we also mapped the dispersal history of GETV inferred until three years in the past: 2000, 2007, and 2015, which allows visualizing the progression of the virus spread.

Supplementary Figures

Figure S1. Continuous environmental variables tested for their impact on the dispersal of GETV lineages in China.

Figure S2. GETV positive cases bar plots. The bar plot under the x-axis represents the number of reported cases of GETV infected mammals in China since 2015. (A) Number of GETV cases in seven regions of China over three time periods from 2015-2017, 2018-2019 and 2020-2021. (B) Number of GETV cases in each season from 2015 to 2021.

Figure S3. Characterization of GETV in cells and suckling mice. (A) 293T or U251 cells were infected with GETV. At 48 hpi, cytopathic changes were observed. (B) Immunofluorescence of GETV Capside protein monoclonal antibody (green) detected in infected 293T or U251 cells, respectively, (blue is DAPI). All fluorescent images were taken at 20×magnification. (C) Weight of mice after infection with GETV. ICR suckling mice (3-day-old) were infected s.c. with 25 μL of GETV ($TCD_{50}=10^{6.5}$/100 ul) or with DMEM. The weight of the mice is plotted against the time of infection. (D) Survival of mice after infection with GETV. No death was detected after hour 80 PI in DMEM group but all the suckling mice in the infected group died after hour 80 PI. Survival analysis was performed in GraphPad software. The significance between surviva lsofmice infected with GETV and DMEM was estimated using a log rank test; ***$P < 0.001$. (E) Clinic sysptoms of mice after infect of GETV.

Figure S4. Comparison of skygrid viral effective population size reconstruction with time-varied covariates. Each plot depicts the mean effective population size trajectory (dark blue), its corresponding 95% Bayesian credibility interval region (light blue), and a time-varying covariate (dark red). (A)The covariates are: annual forest area, (B) annual precipitation, (C) annual pork production, (D) and annual mean temperature.

Figure S5. Analysis of lineage dispersal events associated with the maximum clade credibility tree obtained from the continuous phylogeographic inference.

Figure 1

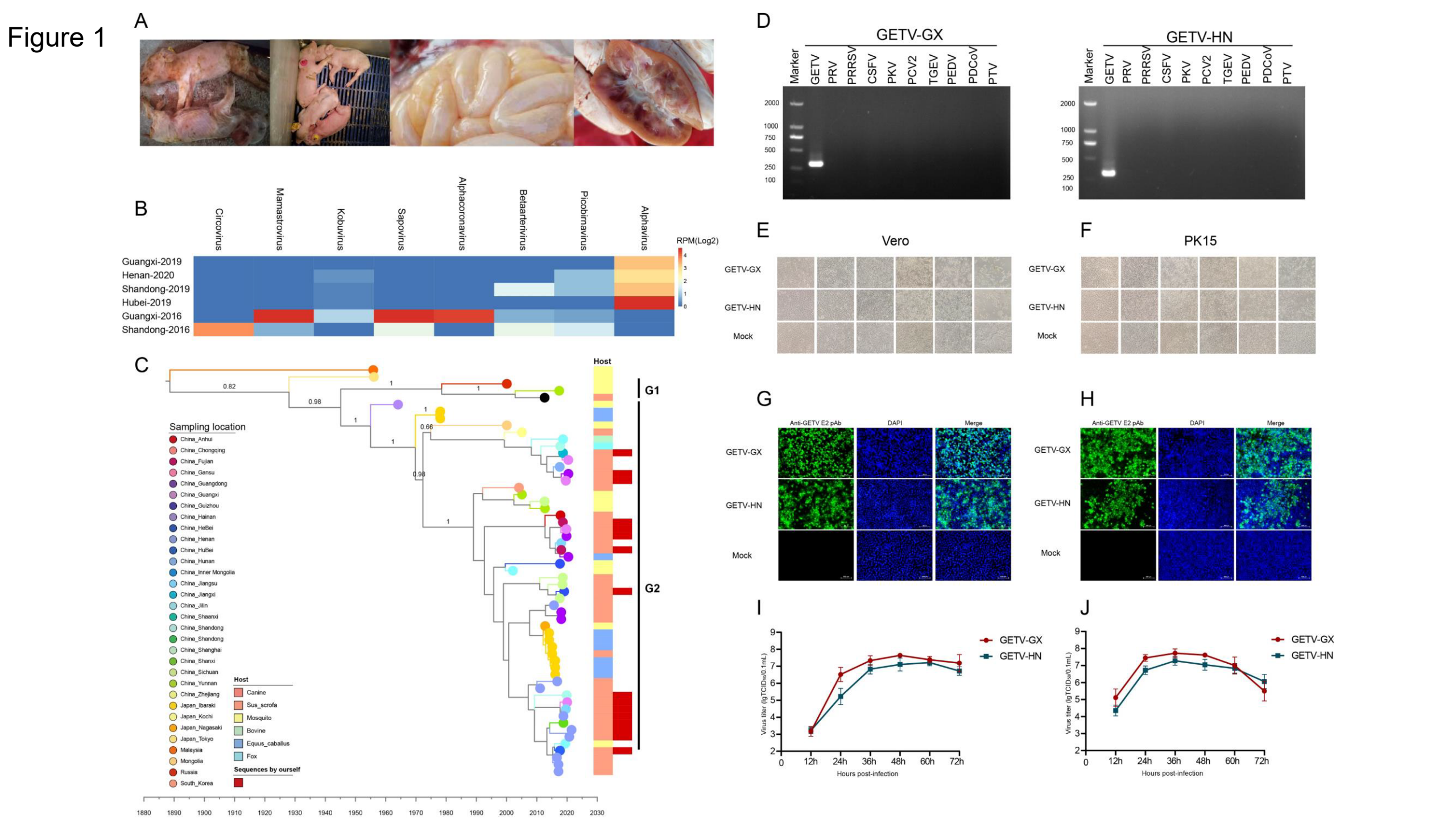

A
B
Circovirus
Mamastrovirus
Kobuvirus
Sapovirus
Alphacoronavirus
Betaarterivirus
Picobirnavirus
Alphavirus
RPM(Log2)
Guangxi-2019
Henan-2020
Shandong-2019
Hubei-2019
Guangxi-2016
Shandong-2016
C
Host
G1
G2
Sampling location
China_Anhui
China_Chongqing
China_Fujian
China_Gansu
China_Guangdong
China_Guangxi
China_Guizhou
China_Hainan
China_HeBei
China_Henan
China_HuBei
China_Hunan
China_Inner Mongolia
China_Jiangsu
China_Jiangxi
China_Jilin
China_Shaanxi
China_Shandong
China_Shandong
China_Shanghai
China_Shanxi
China_Sichuan
China_Yunnan
China_Zhejiang
Japan_Ibaraki
Japan_Kochi
Japan_Nagasaki
Japan_Tokyo
Malaysia
Mongolia
Russia
South_Korea
Host
Canine
Sus_scrofa
Mosquito
Bovine
Equus_caballus
Fox
Sequences by ourself
1880 1890 1900 1910 1920 1930 1940 1950 1960 1970 1980 1990 2000 2010 2020 2030
D
GETV-GX
GETV-HN
Marker
GETV
PRV
PRRSV
CSFV
PKV
PCV2
TGEV
PEDV
PDCoV
PTV
2000
1000
750
500
250
100
E
Vero
F
PK15
GETV-GX
GETV-HN
Mock
G
H
Anti-GETV E2 pAb
DAPI
Merge
I
J
Virus titer (lgTCID50/0.1mL)
Hours post-infection
0 12h 24h 36h 48h 60h 72h

Figure 2

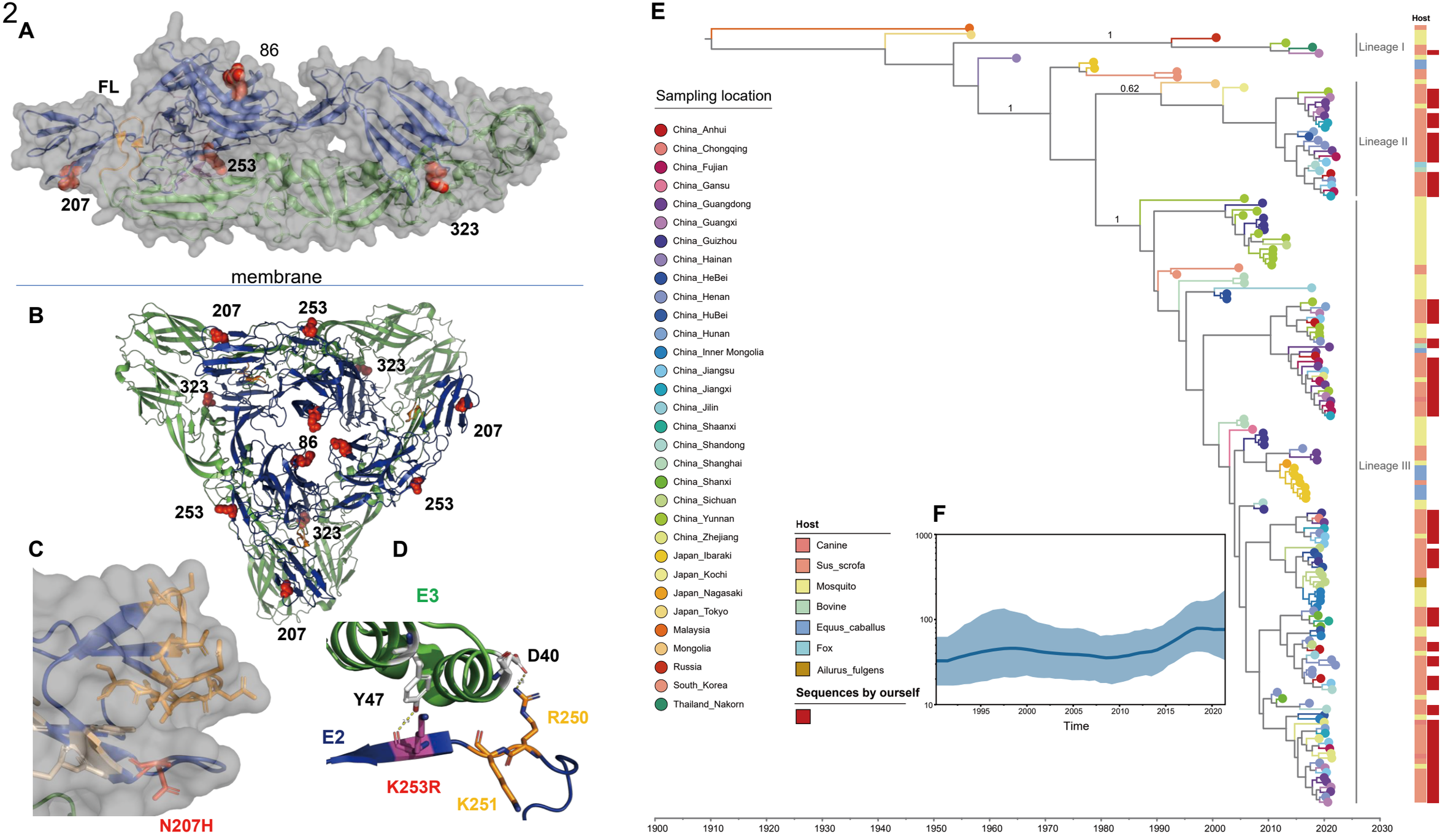

A
86
FL
253
207
323
membrane
B
207
253
323
323
207
86
253
253
323
C
207
N207H
D
E3
D40
Y47
R250
E2
K253R
K251
E
Sampling location
China_Anhui
China_Chongqing
China_Fujian
China_Gansu
China_Guangdong
China_Guangxi
China_Guizhou
China_Hainan
China_HeBei
China_Henan
China_HuBei
China_Hunan
China_Inner Mongolia
China_Jiangsu
China_Jiangxi
China_Jilin
China_Shaanxi
China_Shandong
China_Shanghai
China_Shanxi
China_Sichuan
China_Yunnan
China_Zhejiang
Japan_Ibaraki
Japan_Kochi
Japan_Nagasaki
Japan_Tokyo
Malaysia
Mongolia
Russia
South_Korea
Thailand_Nakorn
1
1
0.62
1
Lineage I
Lineage II
Lineage III
Host
F
Host
Canine
Sus_scrofa
Mosquito
Bovine
Equus_caballus
Fox
Ailurus_fulgens
Sequences by ourself
1000
100
10
1995
2000
2005
2010
2015
2020
Time
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
2020
2030

Figure 3

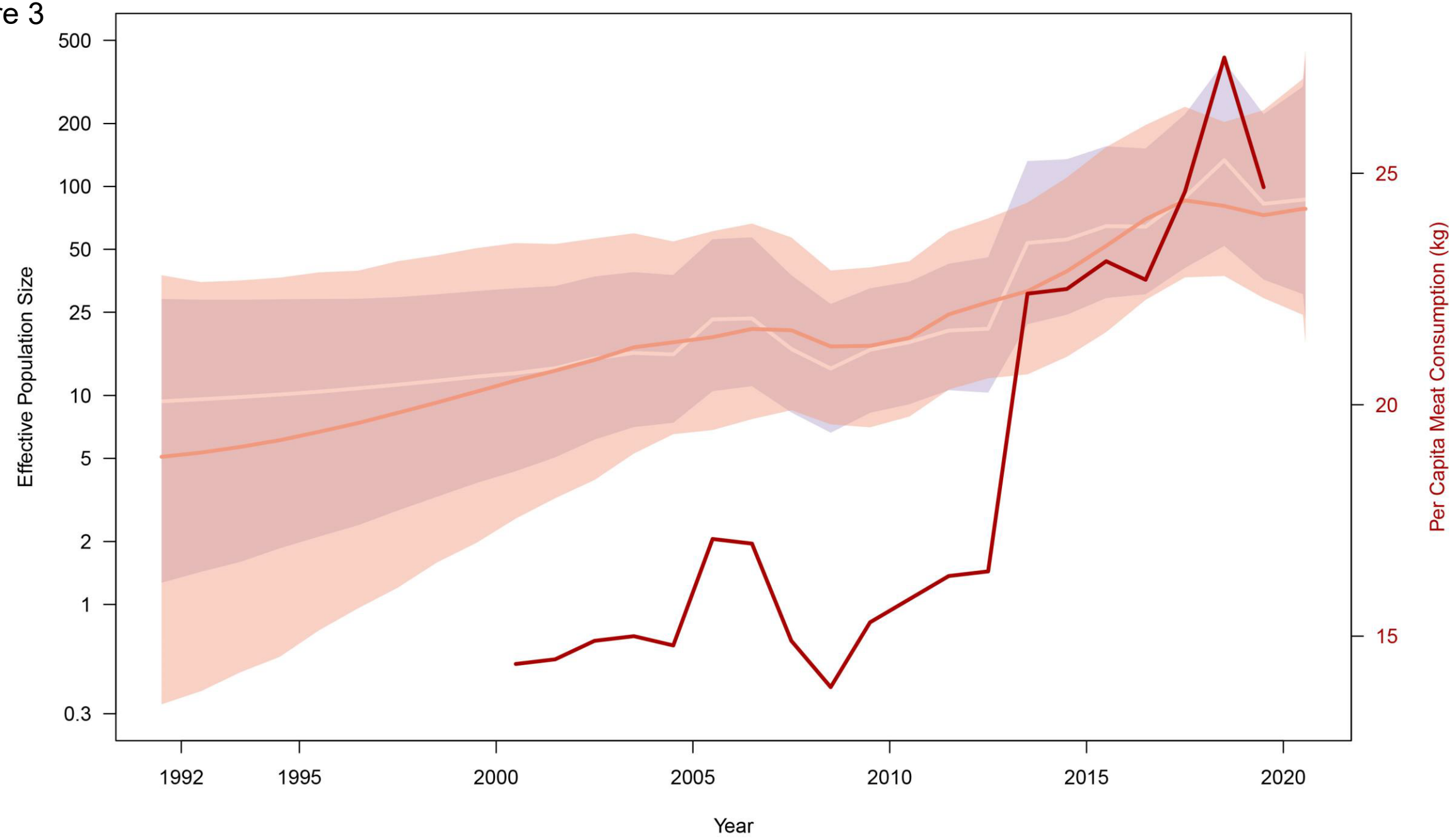

Effective Population Size
500
200
100
50
25
10
5
2
1
0.3
1992
1995
2000
2005
2010
2015
2020
Year
25
20
15
Per Capita Meat Consumption (kg)

Figure 4

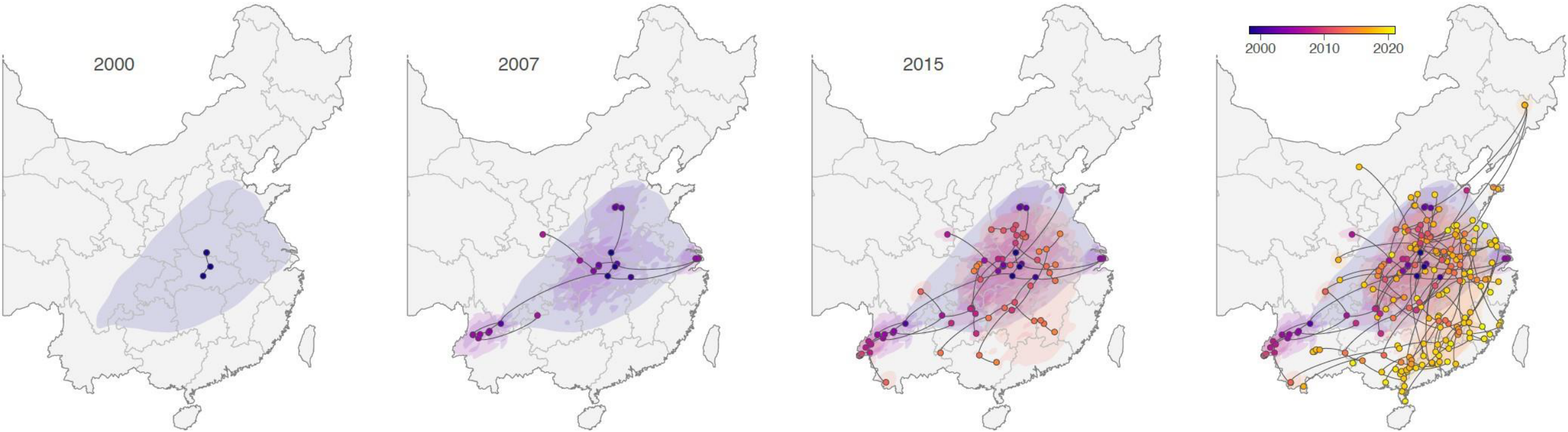


Figure S1

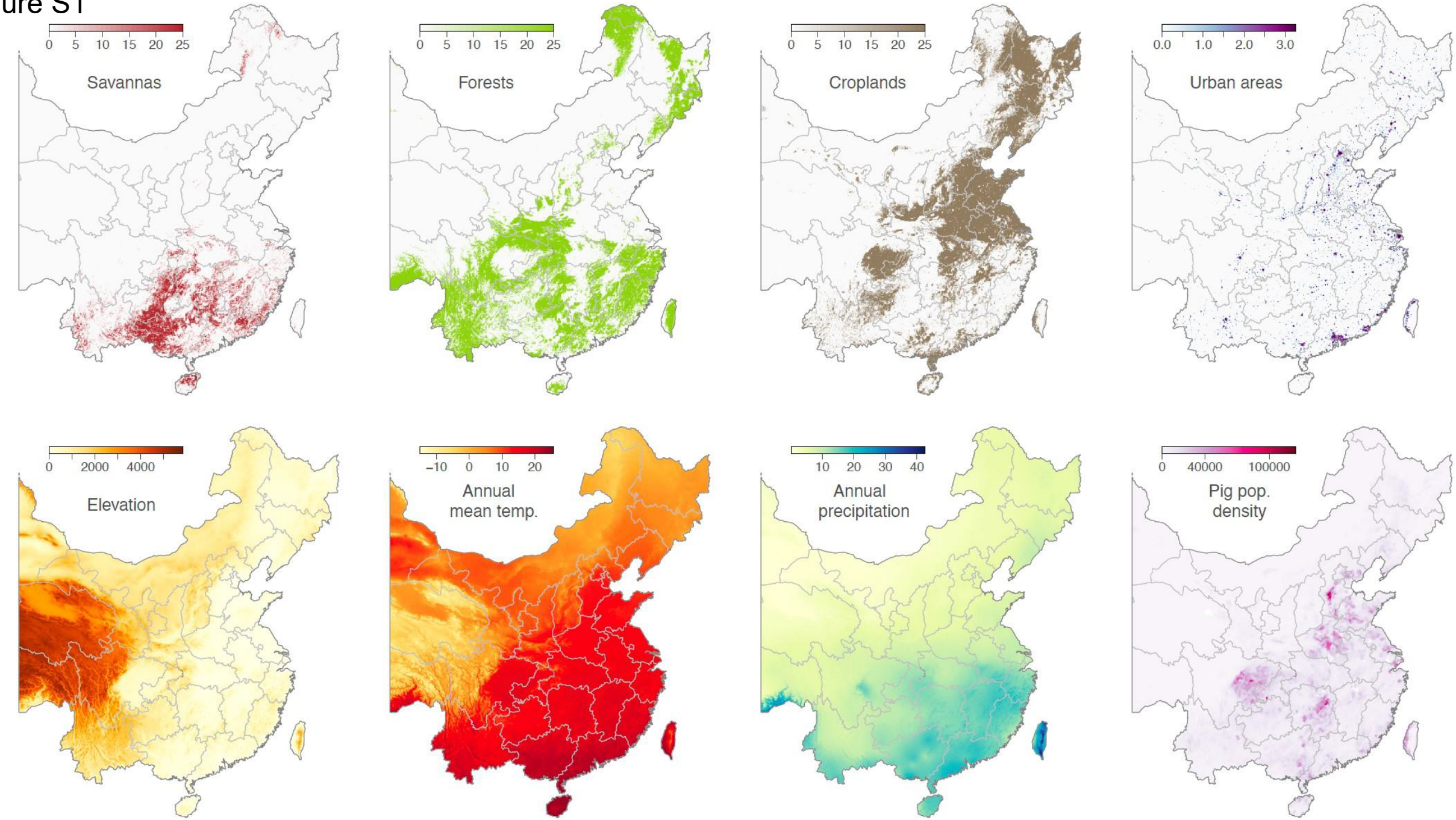

0 5 10 15 20 25
Savannas
0 5 10 15 20 25
Forests
0 5 10 15 20 25
Croplands
0.0 1.0 2.0 3.0
Urban areas
0 2000 4000
Elevation
−10 0 10 20
Annual mean temp.
10 20 30 40
Annual precipitation
0 40000 100000
Pig pop. density

Figure S2

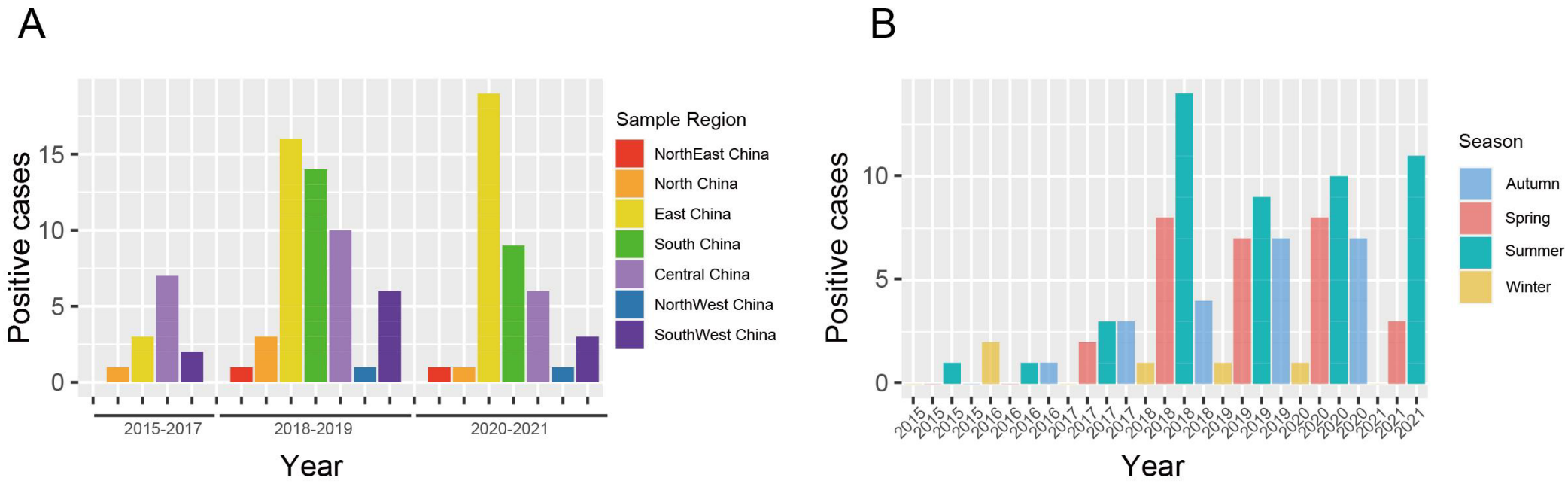

A
Positive cases
15
10
5
0
2015-2017
2018-2019
2020-2021
Year
Sample Region
NorthEast China
North China
East China
South China
Central China
NorthWest China
SouthWest China
B
Positive cases
10
5
0
2015
2016
2017
2018
2019
2020
2021
Year
Season
Autumn
Spring
Summer
Winter

Figure S3

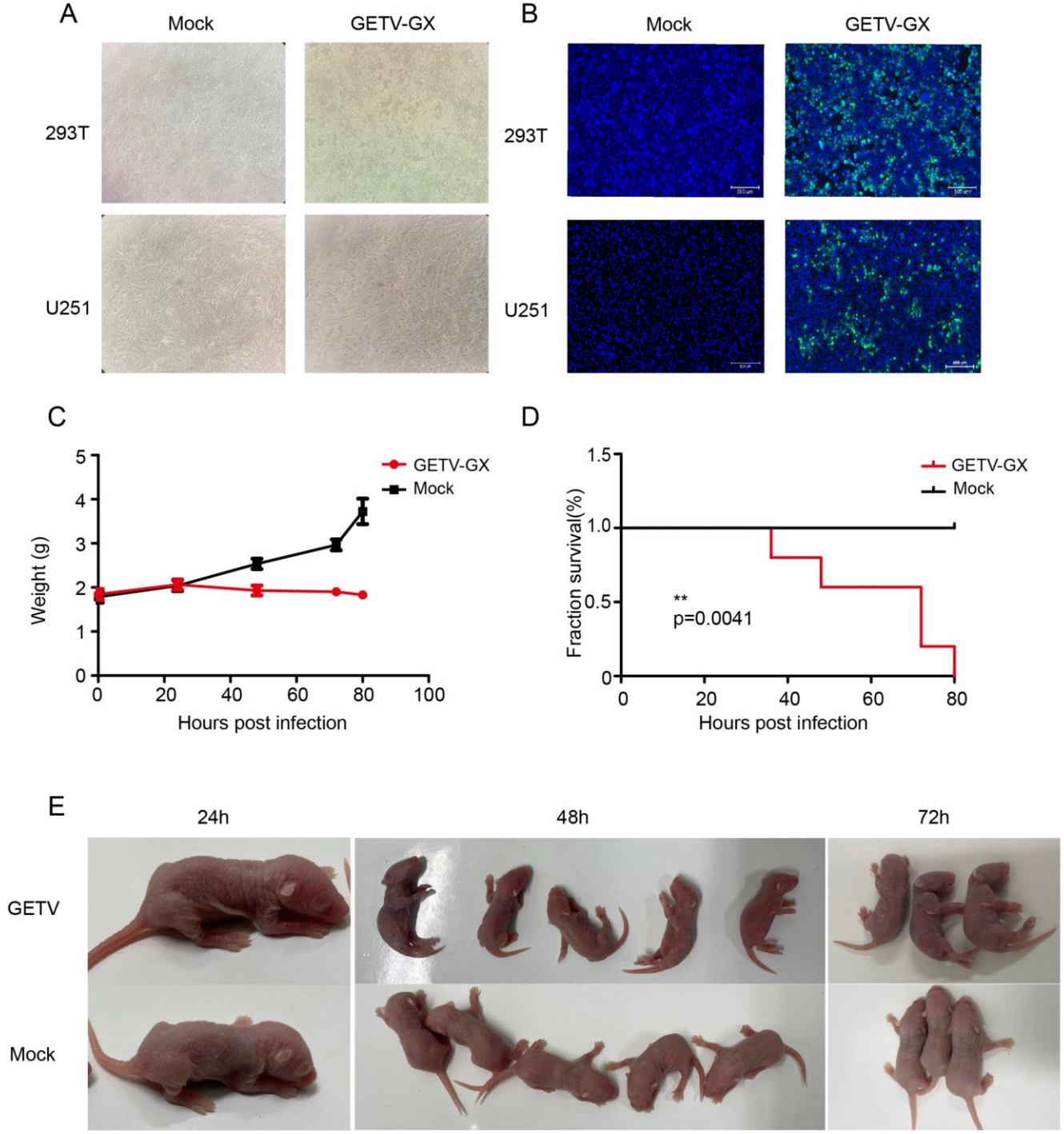

A
Mock
GETV-GX
293T
U251
B
Mock
GETV-GX
293T
U251
C
GETV-GX
Mock
Weight (g)
Hours post infection
D
GETV-GX
Mock
Fraction survival(%)
**
p=0.0041
Hours post infection
E
24h
48h
72h
GETV
Mock

Figure S4

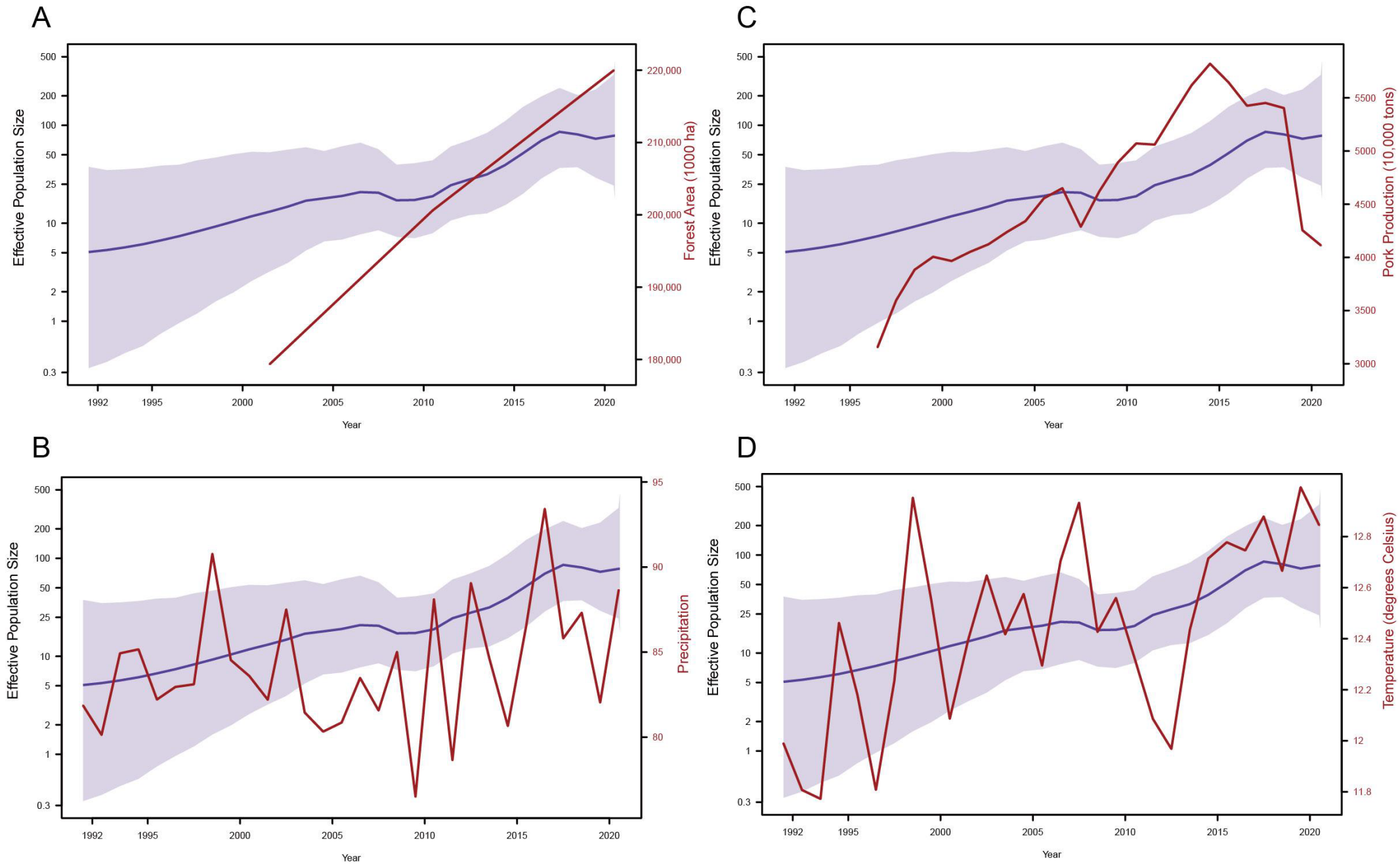
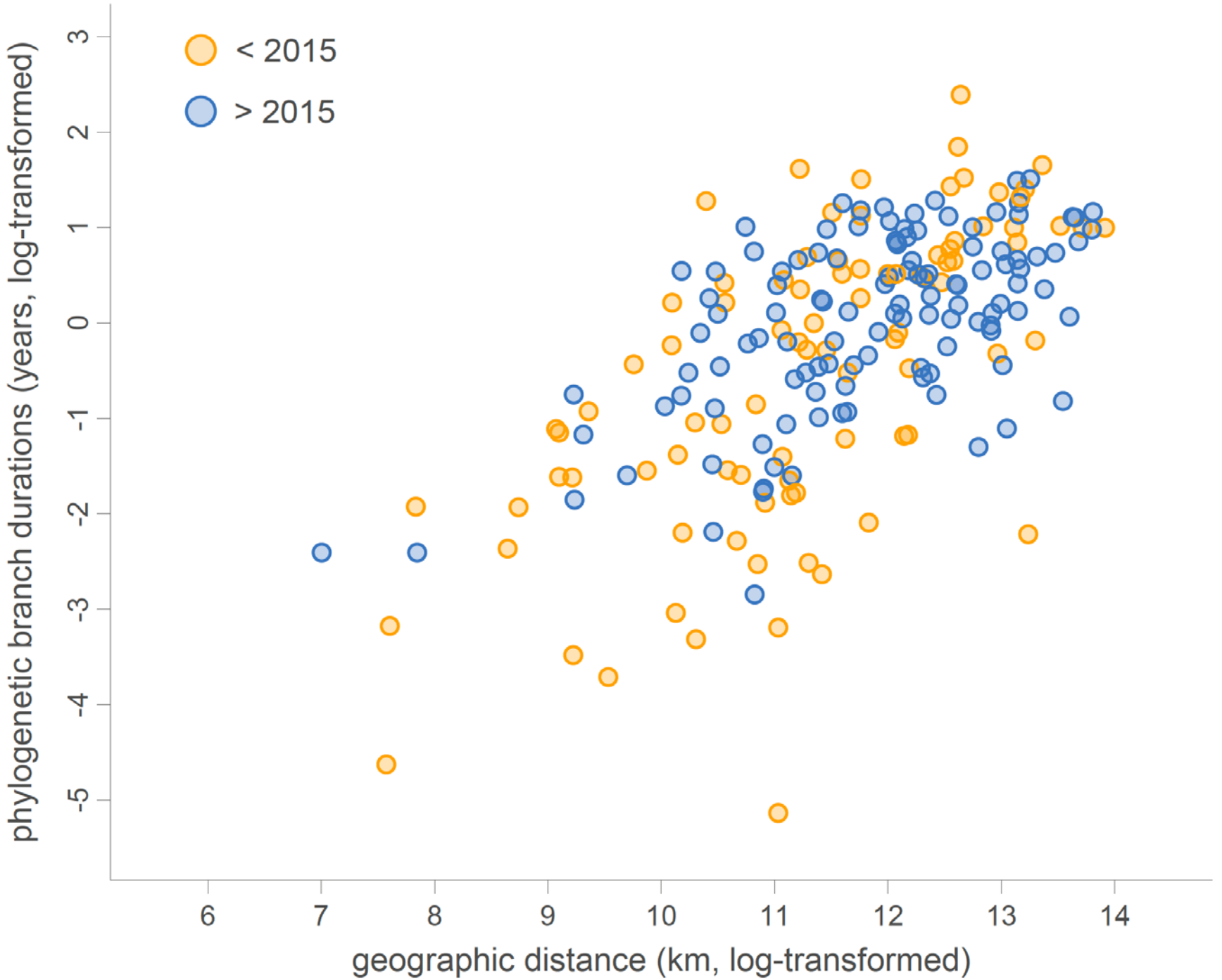

A
Effective Population Size
500
200
100
50
25
10
5
2
1
0.3
1992
1995
2000
2005
2010
2015
2020
Year
220,000
210,000
200,000
190,000
180,000
Forest Area (1000 ha)
C
Effective Population Size
Year
5500
5000
4500
4000
3500
3000
Pork Production (10,000 tons)
B
Effective Population Size
Year
95
90
85
80
Precipitation
D
Effective Population Size
Year
12.8
12.6
12.4
12.2
12
11.8
Temperature (degrees Celsius)

Figure S5


< 2015
> 2015
phylogenetic branch durations (years, log-transformed)
3
2
1
0
-1
-2
-3
-4
-5
6
7
8
9
10
11
12
13
14
geographic distance (km, log-transformed)