Supplementary Materials

Materials and Methods

1. Biological characterization

1.1 Virus titration

Virus titers of the two strains were measured, expressed as a half tissue culture infection dose ($TCID_{50}$). In a nutshell, a 10-fold gradient dilution of virus suspension was inoculated into Vero cells cultured in 96-well plates with a total of 11 gradients diluted, and each gradient had 8 replicates. The cytopathic effect (CPE) was observed after 72h incubation at 37°C, and $TCID_{50}$ was calculated using Reed-Muench method (1).

1.2. One-step growth curve determination

The one-step growth curves of two GETV isolates were measured on Vero cells and PK15 cells, respectively. The multiplicity of infection (MOI) was calculated according to the measured $TCID_{50}$. Vero cells and PK15 cells were cultured in 24-well plate and inoculated with MOI=0.001 and MOI=0.1, respectively. Cell cultures were collected at 12h, 24h, 36h, 48h, 60h, and 72h post infection (hpi) and their corresponding CPE was observed, $TCID_{50}$ of these cultures' supernatants were measured by Reed-Muench method in turn, with three replicates for each hpi.

1.3. Immunofluorescence assay (IFA)

IFA was performed on monolayer cultured cells in 96-well plates, and GETV isolates were inoculated on Vero cells and PK15 cells according to MOI=0.001 and MOI=0.1, respectively. After incubating for 1h at 37°C, remove the inoculum and replace with fresh DMEM containing 2% FBS to maintain the culture for 24h. The cells were then fixed with 4% paraformaldehyde, incubated at room temperature for 30 min and washed with PBS for 3 times. Next, the cells were sealed with 5% skim milk, incubated at 37°C for 1h and washed three times with PBS. The cells reacted with 1:500 diluted GETV-E2 poly-antibody (prepared in our laboratory) and were incubated at 37°C for 1h, then inoculated with FITC-conjugated secondary antibody at 37°C for 1h. Finally, add 4',6-

diamidino-2-phenylindole (DAPI, Solarbio, China) to stain the nucleus, the cells were incubated for 10 min at room temperature and observed under a differential fluorescence microscope (Nikon, Tokyo, Japan).

1.4. Mouse infection test of GETV

In the pathogenicity experiments on mouse models of GETV-GX strain, 3-day-old ICR suckling mice were inoculated intracranially with 25 μl of $10^{6.5}$ $TCID_{50}$/0.1 ml GETV viral solution or with DMEM. Weight and survival status were observed and recorded every 24 hours until all GETV-infected suckling mice died. Mice injected with DMEM (Mock group) were euthanized. Survival analysis was performed in GraphPad software. The significance between survival mice infected with GETV and DMEM was estimated using a log rank test; ***$P < 0.001$.

2. Investigating the impact of environmental factors on the dispersal location and dispersal velocity of GETV lineages

For investigating the impact of environmental factors on the dispersal location of GETV lineages, we used an analytical procedure where mean environmental values are extracted and compared at node positions in inferred (E_{inferred}) and simulated ($E_{\text{simulated}}$) trees (2). Simulated trees were obtained by re-simulating a relaxed random walk diffusion process along the same trees sampled from the posterior distribution under two constraints: the location inferred for the root of the trees is unchanged, and the simulated tree node positions cannot fall in non-accessible areas (such as sea areas). $E_{\text{simulated}}$ values thus correspond to the distribution of mean environmental values explored by phylogenetic nodes under a null dispersal model, i.e. a dispersal scenario that is not impacted by underlying environmental factors. For each tested environmental factor, the E_{inferred} distribution was compared to the $E_{\text{simulated}}$ one to assess if GETV lineages tended to preferentially circulate within or avoid specific environmental conditions, a comparison formalized by the approximation of a Bayes factor support (2).

For investigating the impact of environmental factors on the dispersal velocity of GETV lineages, we employed an analytical procedure where dispersal durations associated with phylogenetic branches are compared to environmental distances computed on each environmental grid (or "raster"). These environmental distances were computed with two path models: the least-cost path model (3) and the path model implemented in the program Circuitscape (4), the latter using circuit theory to accommodate uncertainty in the route taken. The algorithm implemented in Circuitscape treats a given environmental raster as a conductance (facilitating movement) or as a resistance (impeding movement) factor. In the context of the present study, we tested the potential impact of each environmental factor acting once as a conductance and once as a resistance factor. Moreover, for each environmental factor, we generated several distinct rasters by transforming the original raster cell values with the following formula: $v_t = 1 + k(v_o/v_{max})$, where v_t and v_o are the transformed and original cell values, and v_{max} the maximum cell value recorded in the raster. The rescaling parameter k here allows the definition and testing of different strengths of raster cell conductance or resistance, relative to the conductance/resistance of a cell with a minimum value set to "1", which corresponds to the "null" raster (see below). For each of the three environmental factors, we tested three different values for k: k = 10, 100 and 1000. Correlations between phylogenetic branch durations and environmental distances are estimated through the computation of the statistic Q, which is the difference between the coefficient of determination obtained when branch durations are regressed against environmental distances computed on the environmental raster and the coefficient of determination obtained when branch durations are regressed against environmental distances computed on a "null" raster (i.e. a uniform raster with a value of "1" assigned to all cells). We estimated a Q statistic for each environmental factor (treated a conductance or a resistance factor) and each of the 1,000 trees sampled from the posterior distribution. An environmental factor was only considered as potentially explanatory if both its distribution of regression coefficients and its associated distribution of Q values

were positive (5), i.e. with at least 90% of positive values. In this case, the statistical support associated with the resulting Q distribution was compared with the corresponding null of distribution of Q values obtained when computing environmental distances for phylogenetic branches of trees simulated under the null dispersal model introduced above. Similar to the investigation of the procedure used to explore the impact of environmental factors on the dispersal locations of lineages, the comparisons between inferred and simulated distributions of Q values were formalized by approximating Bayes factor support (6).

Result

Sequence similarity

A total of 78 GETV positive sample were detected in 16 provinces in China, from 2017 until now. Two pet canines, and one bovine were found infected GETV. There was no difference in genomic structure between the 16 new genome sequences and the known GETV virus, and the nucleotide similarity at the genome-level obtained in this study was 96.98%-99.99%, with reference sequences was 93.99%-99.99%. The similarity between E2 genes obtained in this study at the nucleotide level was 94.24%-100%, and compared with the online reference, the similarity at the nucleotide level was 92.98%-100%. The amino acid similarity was 94.24%-100% and 92.98%-100%, respectively. In addition, the similarity between bovine E2 gene and GETV of infected other species at amino acid level was 94.49%-99.75%, which was the most similar to GETV from *Culex tritaeniorhynchus* Giles. In addition, the similarity between canine E2 gene and GETV infected with other species at amino acid level was 93.48%-100%, which was the most similar to GETV from Pigs.

Reference

1. REED LJ, MUENCH H. 1938. A SIMPLE METHOD OF ESTIMATING FIFTY PER CENT ENDPOINTS12. American Journal of Epidemiology 27:493-497.
2. Dellicour S, Troupin C, Jahanbakhsh F, Salama A, Massoudi S, Moghaddam MK, Baele G, Lemey P, Gholami A, Bourhy H. 2019. Using phylogeographic approaches to analyse the dispersal history, velocity and direction of viral lineages — Application to rabies virus spread

in Iran. Molecular Ecology 28:4335-4350.
3. Dijkstra EW. 1959. A note on two problems in connexion with graphs. Numerische Mathematik 1:269-271.
4. McRae BH. 2006. ISOLATION BY RESISTANCE. Evolution 60:1551-1561.
5. Jacquot M, Nomikou K, Palmarini M, Mertens P, Biek R. 2017. Bluetongue virus spread in Europe is a consequence of climatic, landscape and vertebrate host factors as revealed by phylogeographic inference. Proc Biol Sci 284.
6. Dellicour S, Rose R, Faria NR, Vieira LFP, Bourhy H, Gilbert M, Lemey P, Pybus OG. 2017. Using Viral Gene Sequences to Compare and Explain the Heterogeneous Spatial Dynamics of Virus Epidemics. Molecular Biology and Evolution 34:2563-2571.

Supplementary Figures and Tables

Figure S1. Environmental factors 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 corresponds to DAPI). All fluorescent images were taken at 20×magnification. (C) Weight of mice after infection with GETV. 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 80h in DMEM group but all the suckling mice in the infected group died after 80h . (E) Clinic symptoms of mice after being infected by GETV.

Figure S4. Root-to-tip regression analysis performed with the program TempEst (here best-fitting the root by maximizing the coefficient of determination R^2 of the linear regression).

Figure S5. Comparison of skygrid viral effective population size reconstructions with time-varying covariates. Each plot depicts the mean effective population size trajectory (dark blue), its corresponding 95% highest posterior density interval region (light blue), and a time-varying covariate (dark red). (A) The different covariates are: annual forest area, (B) annual precipitation, (C) annual pork production, and (D) annual mean

192 temperature.

193

194 **Figure S6.** Analysis of lineage dispersal events associated with the maximum clade
195 credibility (MCC) tree obtained from the continuous phylogeographic inference.

196

197 **Table S1.** Positive selected amino acid sites of E2 gene of GETV.

Amino acid site	FEL	SLAC	MEME	FUBAR
86	+	-	+	+
253	-	-	-	+
323	+	-	+	+

198

199 **Table S2.** Impact of several environmental factors on the dispersal location of GETV
200 lineages in China. We report Bayes factor (BF) supports for the association between
201 environmental values and tree node locations. The results are based on 100 posterior
202 trees obtained by spatially-explicit phylogeographic inference. Following Kass &
203 Raftery (1995), we consider a BF value >20 as strong support for a significant
204 correlation between the environmental distances and dispersal durations (in bold).

Environmental factor	Tendency of viral lineages to avoid circulating within specific environmental conditions	Tendency of viral lineages to preferentially circulate within specific environmental conditions
Savannas	0.3	3.5
Forests	0.3	2.9
Croplands	0.1	17.2
Urban areas	0.1	13.3
Elevation	**49.0**	0.0
Annual mean temperature	0.0	**>99**
Annual precipitation	0.1	12.3
Pig population density	0.0	**>99**

205

Figure S1

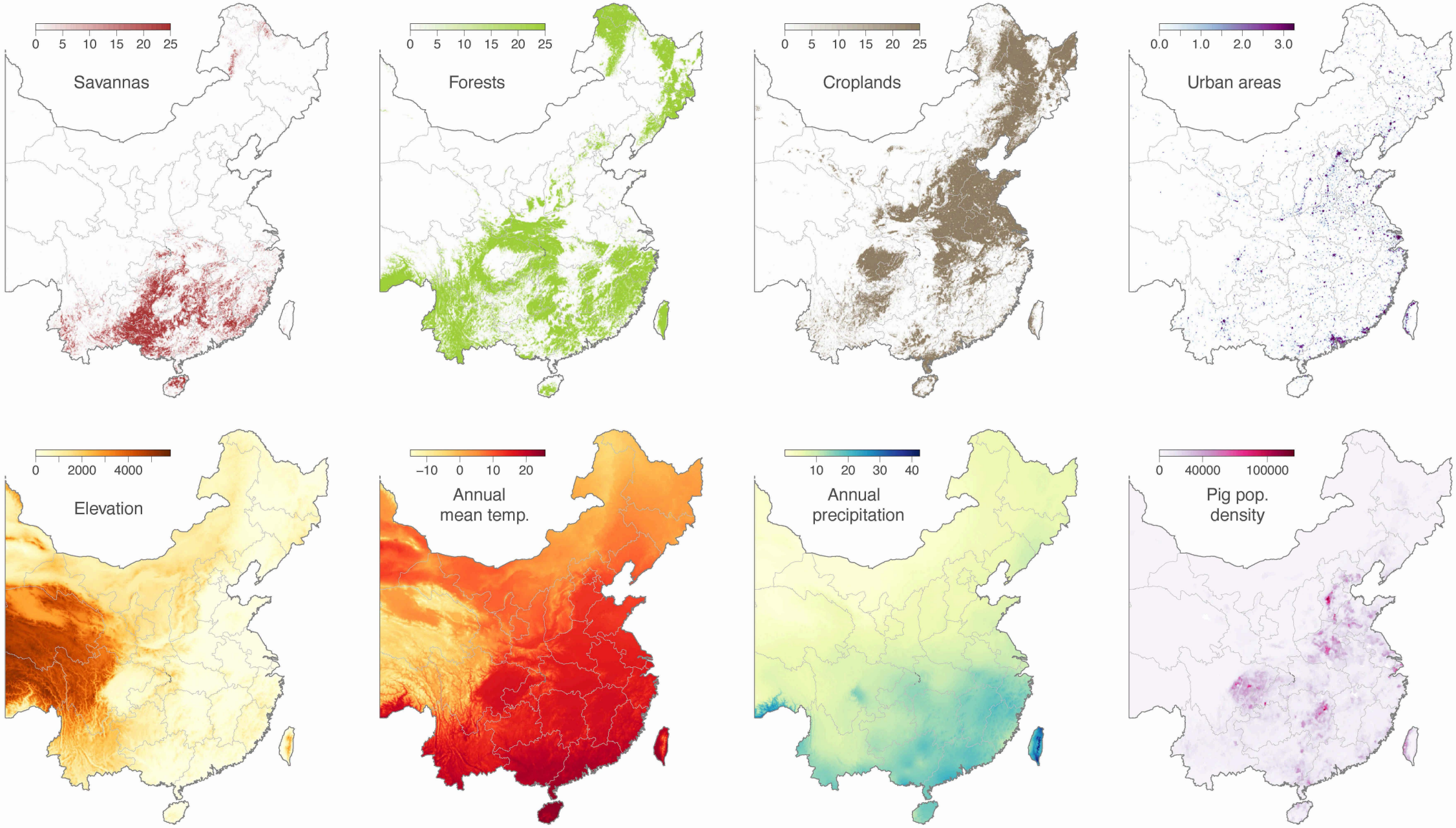

0 5 10 15 20 25
Savannas
0 5 10 15 20 25
Forests
0 5 10 15 20 25
Croplands
0.0 1.0 2.0 3.0
Urban areas
0 2000 4000
Elevation
−10 0 10 20
Annual mean temp.
10 20 30 40
Annual precipitation
0 40000 100000
Pig pop. density

Figure S2

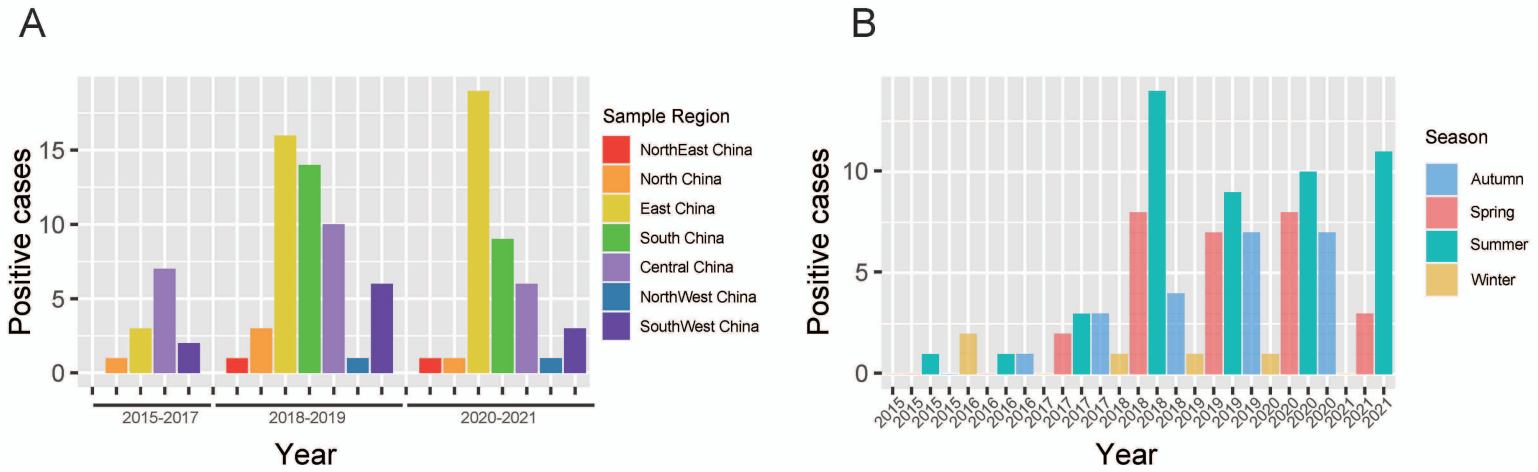

A
Positive cases
0
5
10
15
2015-2017
2018-2019
2020-2021
Year
Sample Region
NorthEast China
North China
East China
South China
Central China
NorthWest China
SouthWest China
B
Positive cases
0
5
10
Year
Season
Autumn
Spring
Summer
Winter

Figure S3

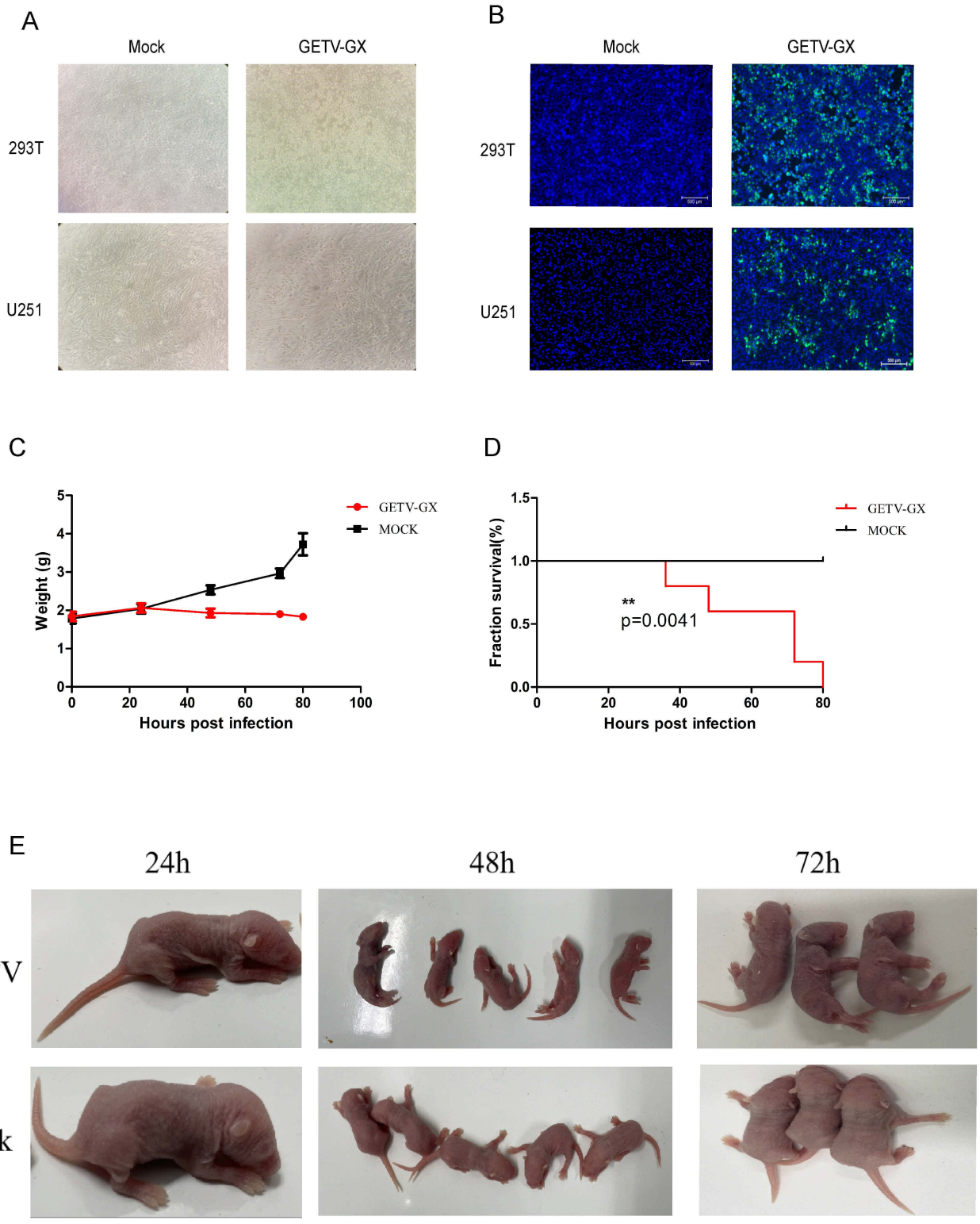

A
Mock
GETV-GX
293T
U251
B
Mock
GETV-GX
293T
U251
C
GETV-GX
MOCK
Weight (g)
Hours post infection
D
GETV-GX
MOCK
Fraction survival(%)
**
p=0.0041
Hours post infection
E
24h
48h
72h
GETV
Mock

Figure S4

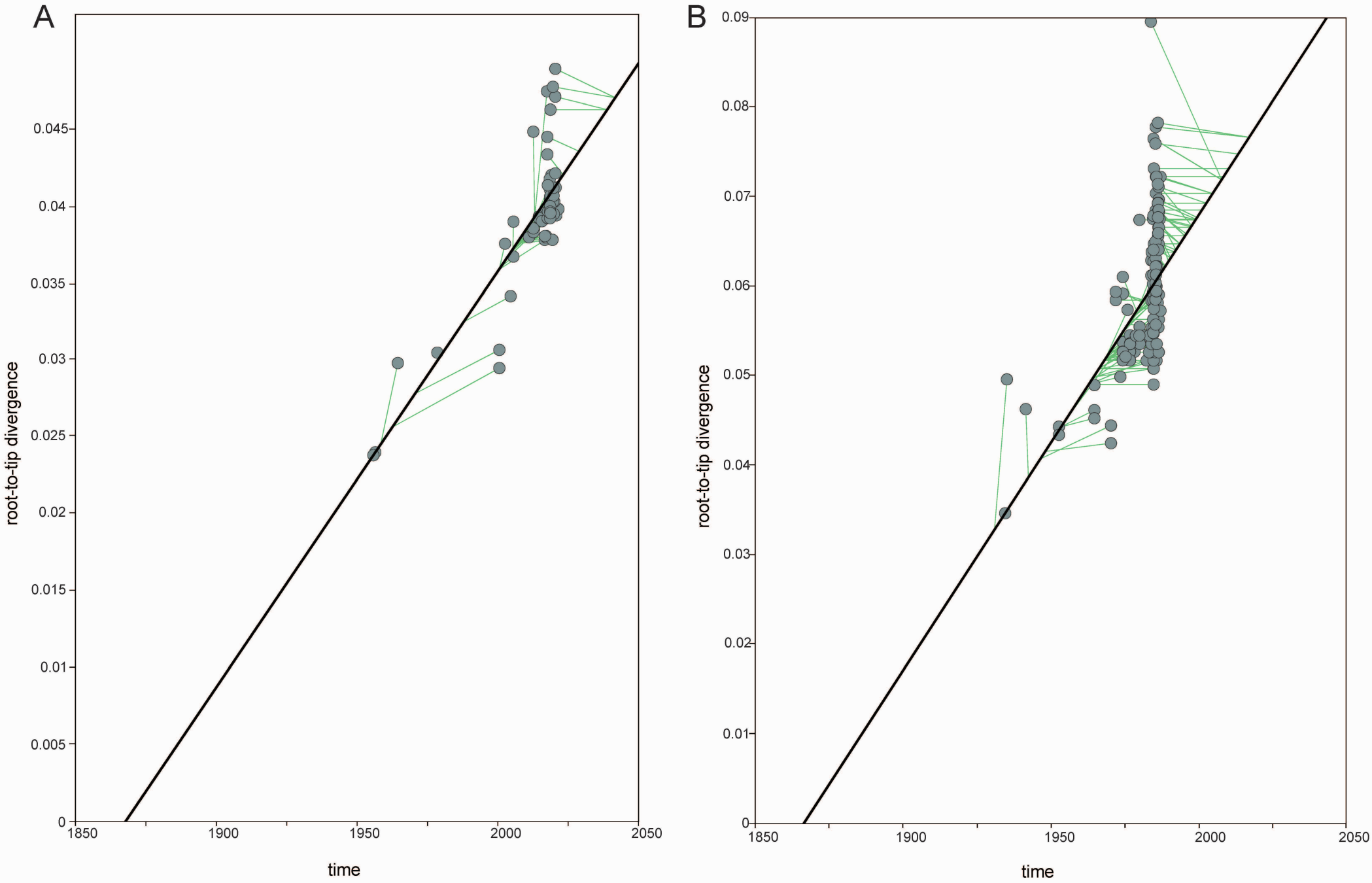

A
B
root-to-tip divergence
time
0
0.005
0.01
0.015
0.02
0.025
0.03
0.035
0.04
0.045
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
1850
1900
1950
2000
2050

Figure S5

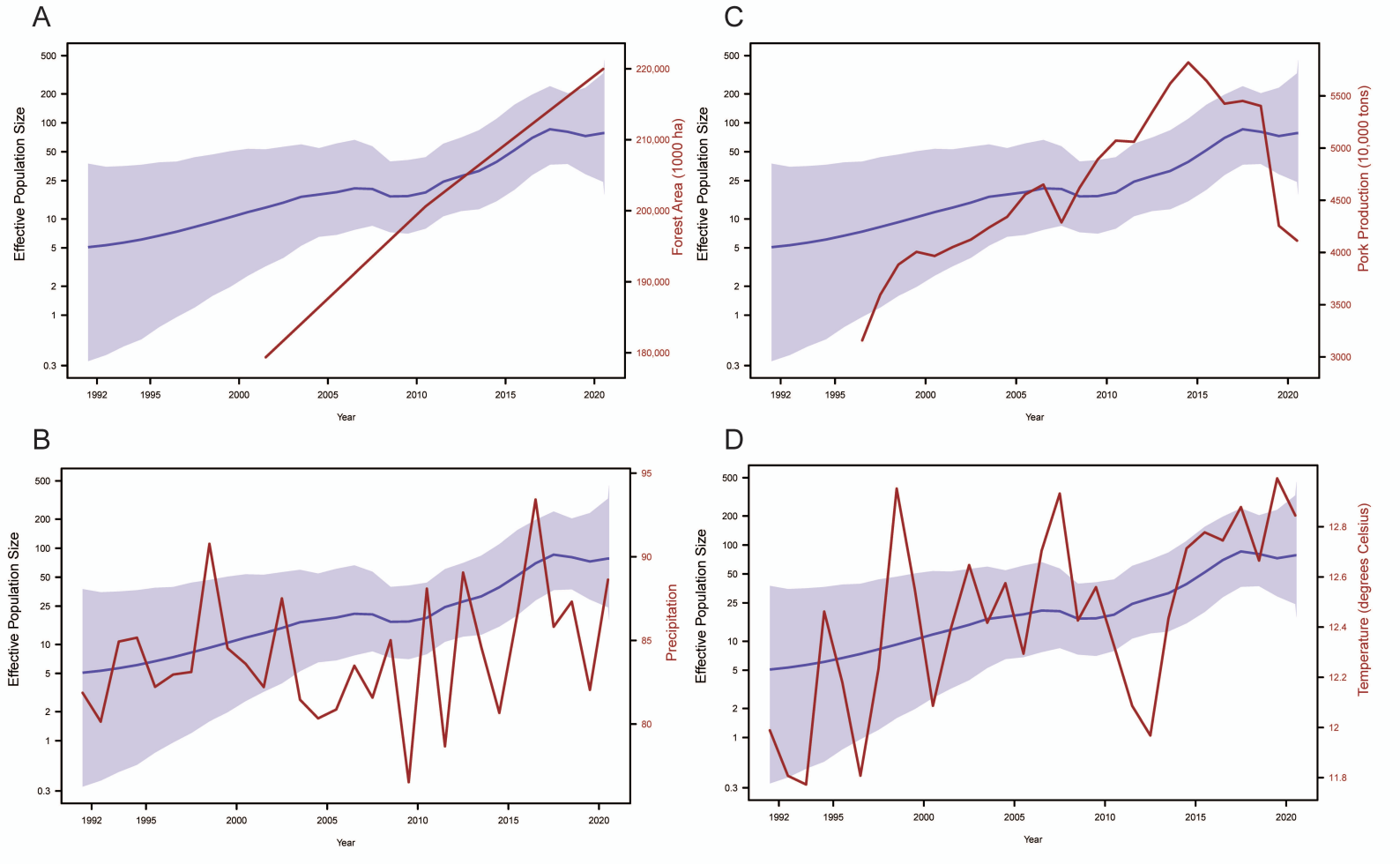
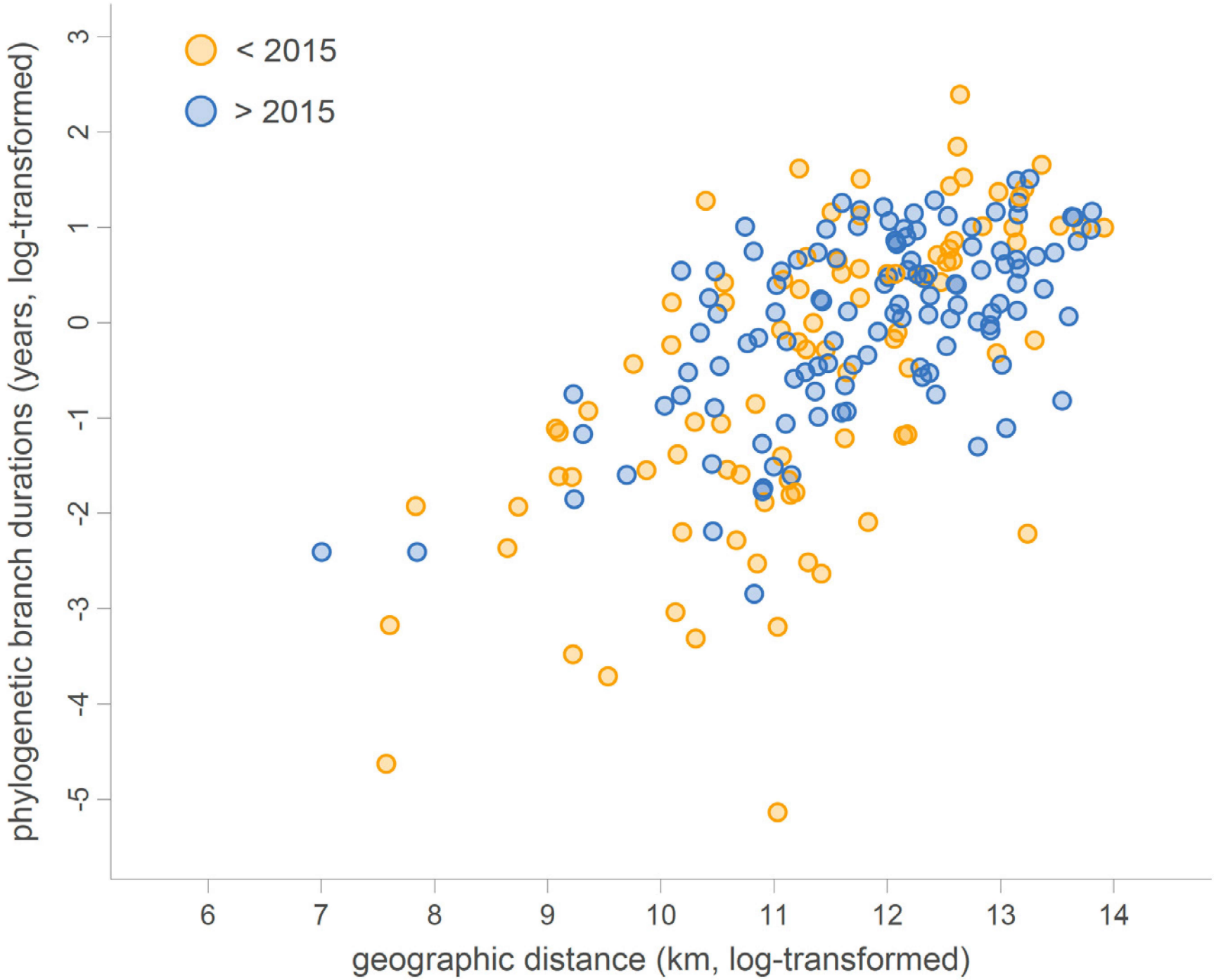

A
Effective Population Size
500
200
100
50
25
10
5
2
1
0.3
1992
1995
2000
2005
2010
2015
2020
Year
Forest Area (1000 ha)
220,000
210,000
200,000
190,000
180,000
C
Effective Population Size
Pork Production (10,000 tons)
5500
5000
4500
4000
3500
3000
B
Precipitation
95
90
85
80
D
Temperature (degrees Celsius)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

Figure S6


< 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)