PNAS
www.pnas.org

Supplementary Information for

Historical and genomic data reveal the influencing factors on global transmission velocity of plague during the Third Pandemic

Lei Xu, Leif Chr. Stige, Herwig Leirs, Simon Neerinckx, Kenneth L. Gage, Ruifu Yang, Qiyong Liu, Barbara Bramanti, Katharine R. Dean, Hui Tang, Zhe Sun, Nils Chr. Stenseth, Zhibin Zhang

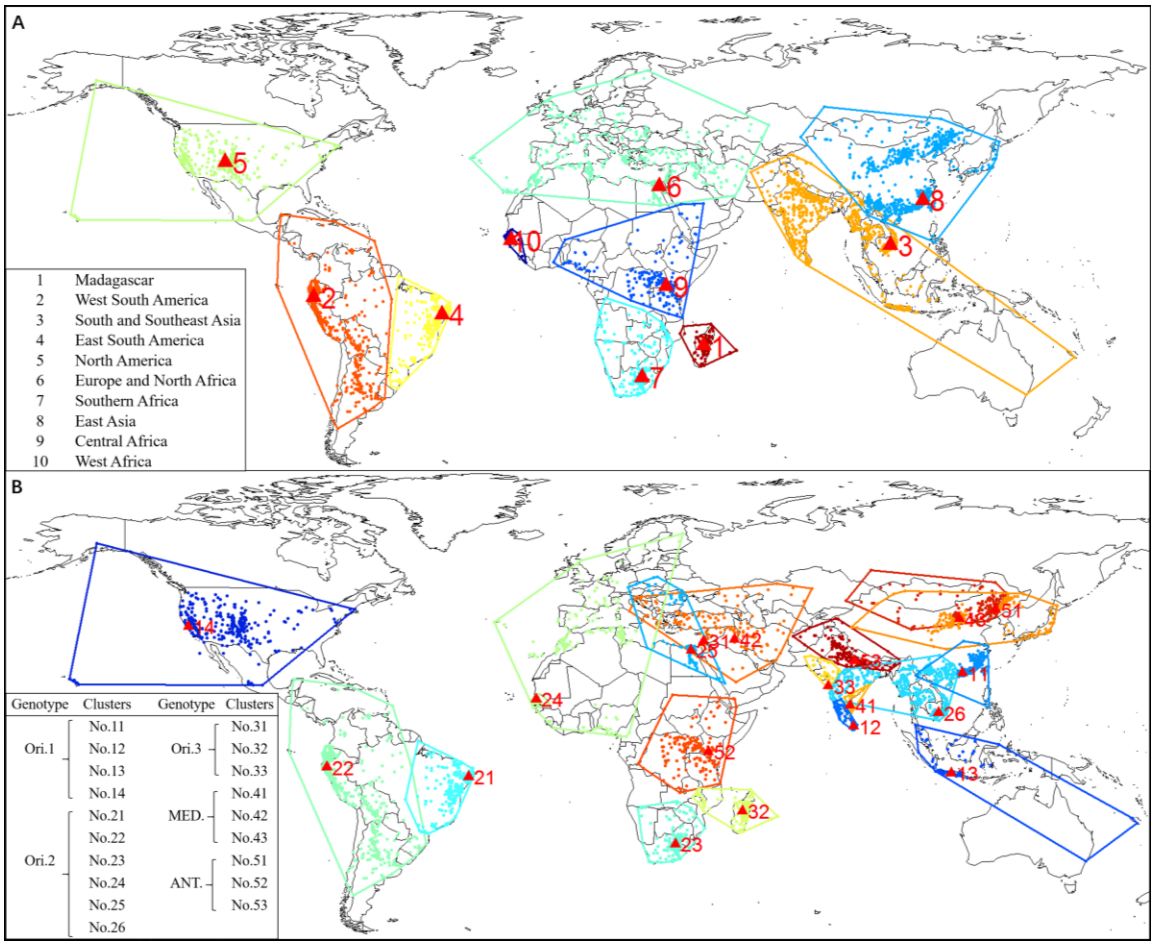
Paste corresponding author name here
Email: zhangzb@ioz.ac.cn or n.c.stenseth@ibv.uio.no

This PDF file includes:

Figs. S1 to S5
Tables S1
References for SI reference citations

www.pnas.org/cgi/doi/10.1073/pnas.1901366116

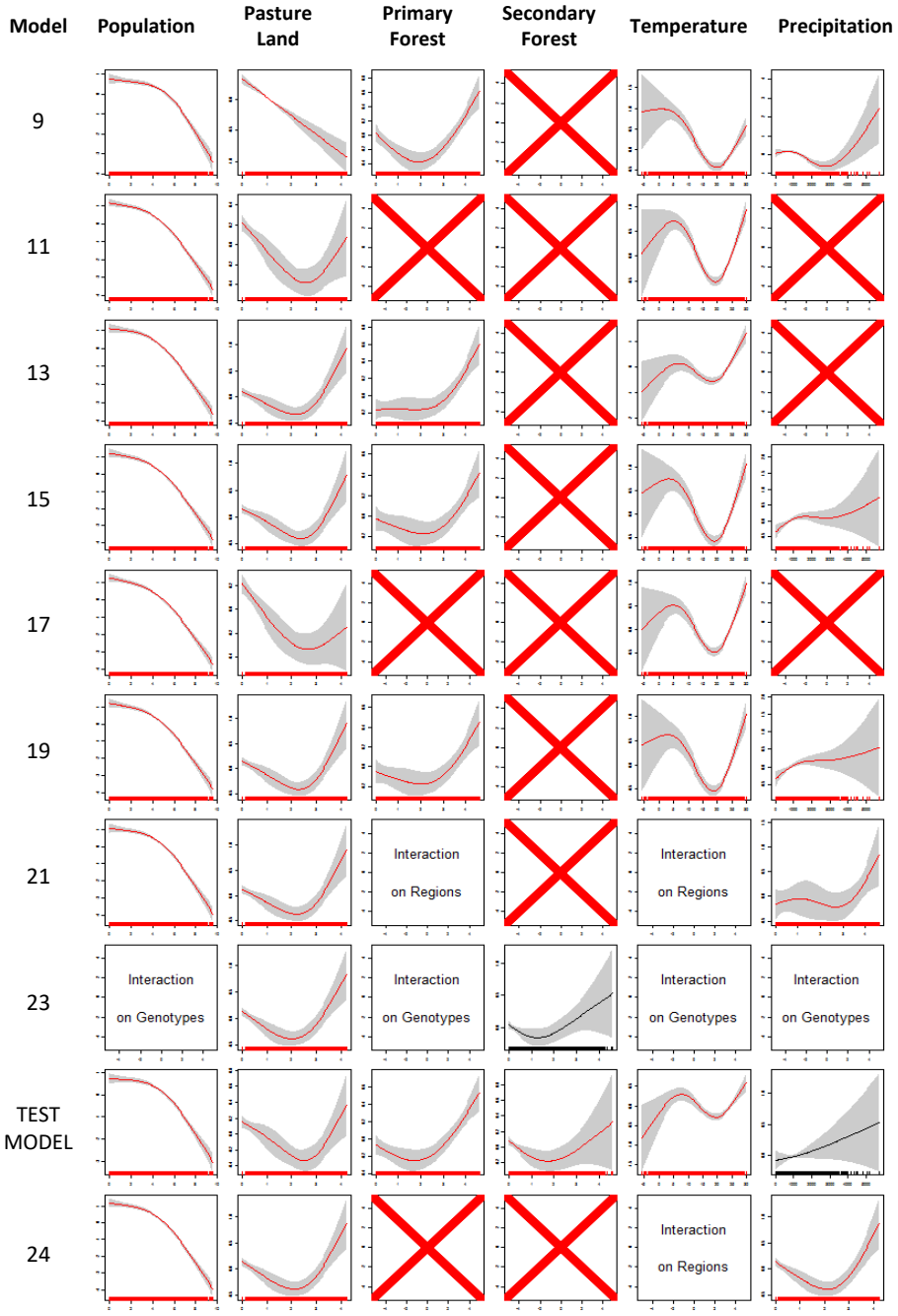
Fig. S1.



Spatial clustering analysis. Global plague infected localities (e.g., counties or villages) were categorized into groups by two methods. **A.** Spatial cluster results based on the density peak method (*1*), using data of plague infected localities (pooled data across plague genotypes). It was used to identify spatial clusters based on the spatial density of plague infected localities, with cluster centers characterized by higher density than neighboring locations and a relatively large distance from locations with higher densities. Red triangles indicate cluster centers and polygons indicate the convex hull of each cluster. The identified clusters were used as "Region" variable in the statistical analysis. **B.** Spatial cluster results based on the density peak method constrained by genotypes, with separate clustering for each plague genotype. Preliminary analyses suggested that plague spread was more parsimoniously modelled as additive functions of Region and Genotype than by a region variable constructed from the genotype-constrained cluster analysis. Convex hulls of clusters 24, 25 and 42 overlapped in south-middle Europe, clusters 51 and 43 overlapped

in North Asia, clusters 11 and 26 overlapped in India and Southeast China, reflecting the existence of multiple genotypes in these regions. Nowadays, the distribution of human plague cases is closely linked to the regions of natural plague foci of the world (*2-4*).

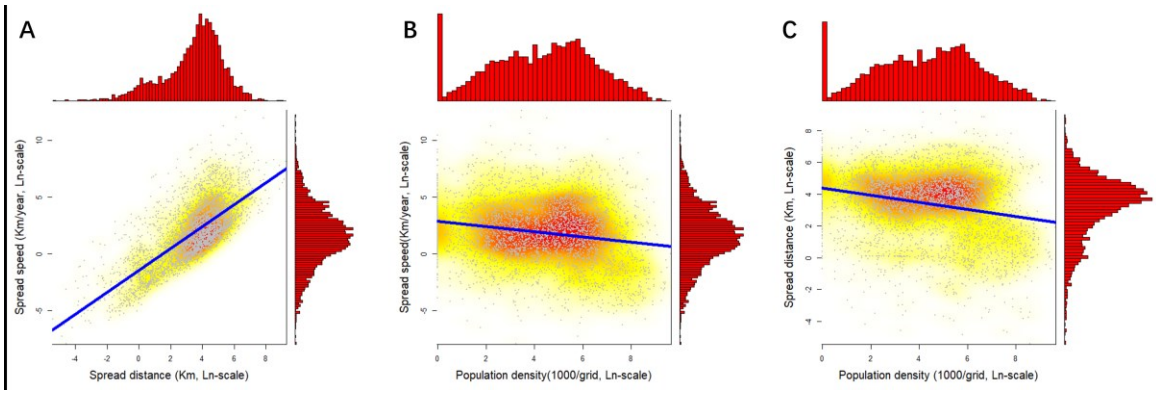
Fig. S2.



Estimated effects of environmental variables on plague spread velocity in alternative model formulations. See Table S1 for equations and out-of-sample predictive power of the different models (identified by Model ID). Test model: spatial and temporal trends were accounted for by using residuals from Base Model 1 as response variable. Red cross: variable not included in model. The results show that estimated effects of population density and temperature were consistent among models. Effects of pasture land, primary forest and precipitation were generally consistent among models that included year and region effects. The precipitation effect was non-significant if regional differences were modelled as a 2-D smooth function of latitude and longitude (Model 13, Test model) rather

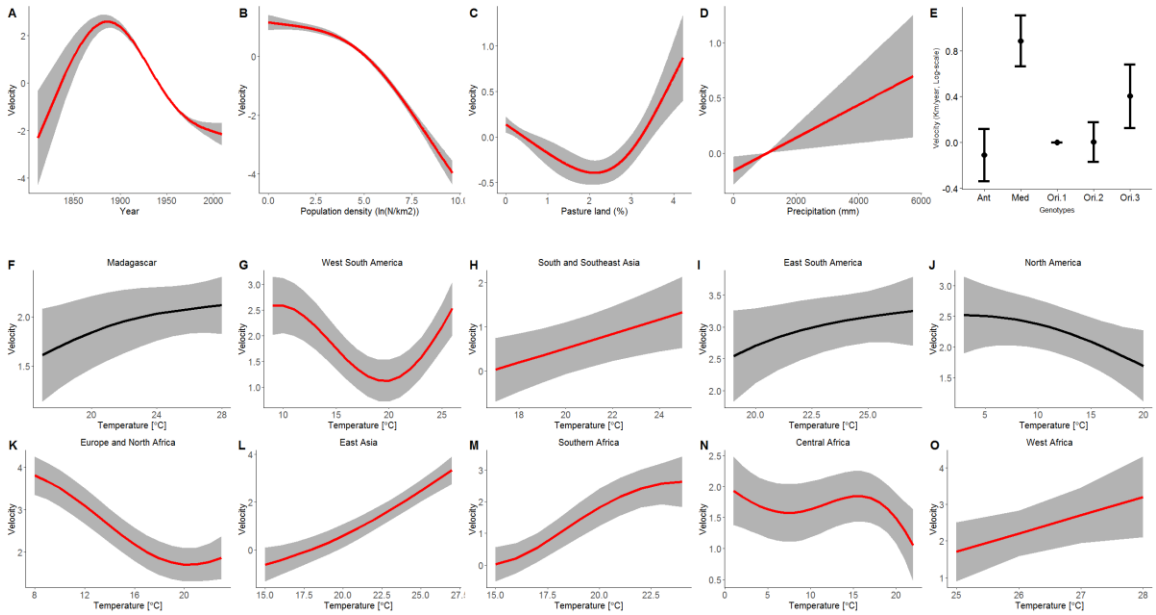
than as a function of the categorical region variable (Models 15, 19, 21, 23, 24). Model 24 is simplified from Model 21 by only including one interaction effect and is also shown in Fig. S4.

Fig. S3.



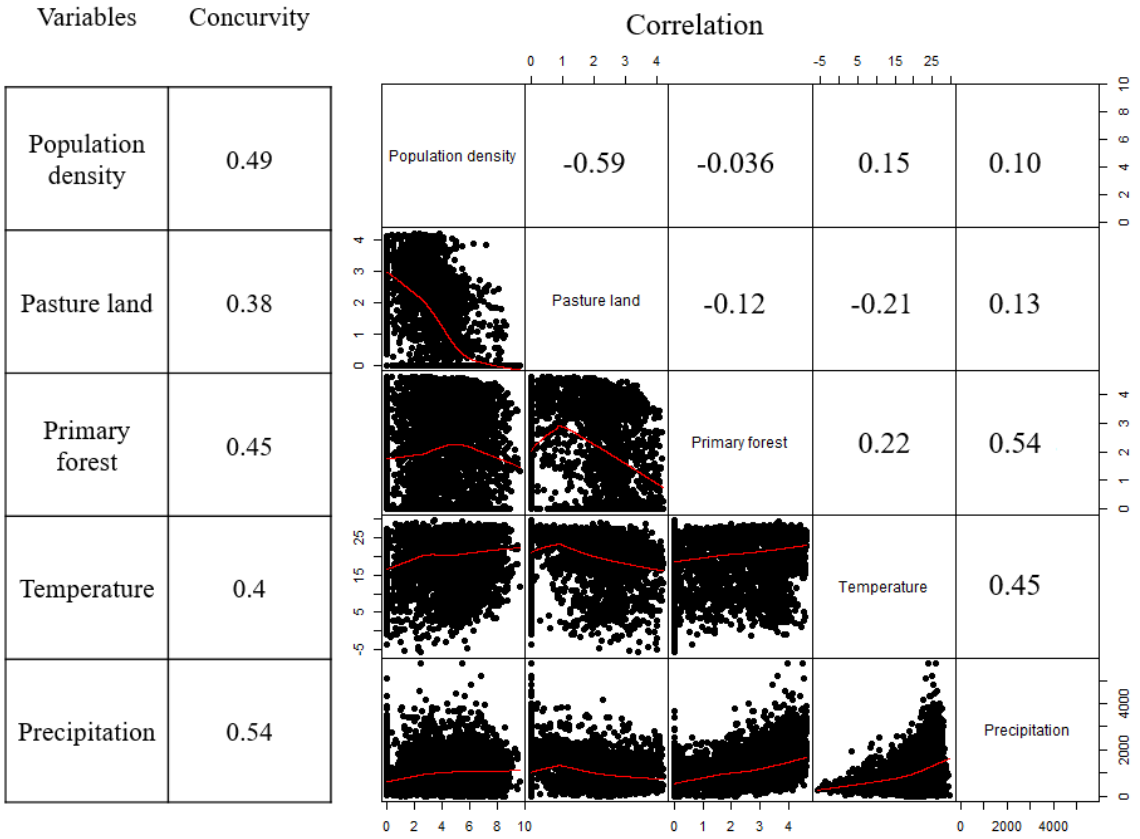
Correlation plots with histogram of pairwise associations among human population density (1000/grid, ln-transformed), plague spread speed (km/year, ln-transformed) and plague spread distance (km, ln-transformed). **A.** Plot between plague spread speed and spread distance. **B.** Plot between spread speed and population density. **C.** Plot between spread distance and population density.

Fig. S4.



Plot of Model 24 from Table S1 showing the region-specific effect of temperature on spread velocity of plague. Each panel shows the partial effect of one variable on plague spread velocity (ln[km/year]). Grey bands (panels A−D, F−O) and bars (panel E) are 95 % nominal confidence bands. Red lines are statistically significant at nominal $p < 0.05$, black lines not. Panel E shows the effects of the categorical variable Genotype, with *ORI.1* as reference level ($p < 0.05$). Panels F−O show interaction effects of region and temperature (with colors reflecting the significance of the temperature effect in each region).

Fig. S5.



Correlation between the environmental predictor variables in the final Generalized Additive Model (GAM). Left-side panels tabulate the concurvity of each variable. The concurvity measures to which degree each smooth term in the GAM can be approximated by one or more of the other smooth terms in the model, and can be viewed as a generalization of multicollinearity. The concurvity was measured by the 'concurvity' function in the mgcv library of R and scales from 0 to 1, with 0 indicating no problem, and 1 indicating total lack of identifiability. Right-side panels show pairwise associations between variables, with nonlinear trendlines shown in red in the lower triangle and the linear product-moment correlation coefficients tabulated in the upper triangle.

Table S1. Model selection. Total degrees of freedom, *d.f.*, refers to predictor functions, with complex models having high *d.f.* CV: out-of-sample root mean squared prediction error calculated by cross validation, iteratively leaving out and making out-of-sample predictions for data for blocks of ten years at a time. Low CV means high predictive power. Models 18, 19, 21, 23, 24 have similarly high predictive power (all CVs < 2.32), but Model 19 (marked by *) is the most parsimonious with smallest *d.f.* (least complex). Therefore, we presented the model results of Model 19 in Fig. 2. To show the region-dependent effect of temperature on spread velocity, we presented the results of Model 24 in Fig. S3. Model 24 simplifies Model 21 by only including one interaction effect.

ID	Model	Formula	Total *d.f.*	CV
1	Base.Model.1	$V_{i.t} = a + b(Year_t) + c(Lon_i.Lat_i) + \varepsilon_{i.t}$	32.81	2.471
2	Base.Model.2	$V_{i.t} = a + b(Year_t) + \acute{d}(Region_i) + \varepsilon_{i.t}$	12.98	2.492
3	Base.Model.3	$V_{i.t} = a + b(Year_t) + d(Genotype_i) + \varepsilon_{i.t}$	7.98	2.498
4	Base.Model.4	$V_{i.t} = a + b(Year_t) + c(Lon_i.Lat_i) + d(Genotype_i) + \varepsilon_{i.t}$	36.98	2.469
5	Base.Model.5	$V_{i.t} = a + b(Year_t) + \acute{d}(Region_i) + d(Genotype_i) + \varepsilon_{i.t}$	16.99	2.480
6	Base.Model.7	$V_{i.t} = a + b(Year_t \times Genotype_i) + d(Genotype_i) + \varepsilon_{i.t}$	15.92	2.503
7	Base.Model.8	$V_{i.t} = a + b(Year_t \times Genotype_i) + d(Genotype_i) + \acute{d}(Region_i) + \varepsilon_{i.t}$	25.26	2.494
8	Env.Model.1	$V_{i.t} = a + e(Pop_{i.t}) + f(Pasture_{i.t}) + h(PriFor_{i.t}) + k(SecFor_{i.t}) + m(Temp_{i.t}) + n(Prec_{i.t}) + \varepsilon_{i.t}$	14.50	2.554
9	Env.Model.1.Red	$V_{i.t} = a + e(Pop_{i.t}) + f(Pasture_{i.t}) + k(SecFor_{i.t}) + m(Temp_{i.t}) + n(Prec_{i.t}) + \varepsilon_{i.t}$	13.19	2.552
10	Env.Model.2	$V_{i.t} = a + b(Year_t) + e(Pop_{i.t}) + f(Pasture_{i.t}) + h(PriFor_{i.t}) + k(SecFor_{i.t}) + m(Temp_{i.t}) + n(Prec_{i.t}) + \varepsilon_{i.t}$	17.63	2.350
11	Env.Model.2.Red	$V_{i.t} = a + b(Year_t) + e(Pop_{i.t}) + f(Pasture_{i.t}) + m(Temp_{i.t}) + \varepsilon_{i.t}$	12.23	2.346
12	Env.Model.3	$V_{i.t} = a + b(Year_t) + c(Lon_i.Lat_i) + e(Pop_{i.t}) + f(Pasture_{i.t}) + h(PriFor_{i.t}) + k(SecFor_{i.t}) + m(Temp_{i.t}) + n(Prec_{i.t}) + \varepsilon_{i.t}$	39.19	2.331
13	Env.Model.3.Red	$V_{i.t} = a + b(Year_t) + c(Lon_i.Lat_i) + e(Pop_{i.t}) + f(Pasture_{i.t}) + h(PriFor_{i.t}) + m(Temp_{i.t}) + \varepsilon_{i.t}$	36.61	2.329
14	Env.Model.4	$V_{i.t} = a + b(Year_t) + \acute{d}(Region_i) + e(Pop_{i.t}) + f(Pasture_{i.t}) + h(PriFor_{i.t}) + k(SecFor_{i.t}) + m(Temp_{i.t}) + n(Prec_{i.t}) + \varepsilon_{i.t}$	28.51	2.326
15	Env.Model.4.Red	$V_{i.t} = a + b(Year_t) + \acute{d}(Region_i) + e(Pop_{i.t}) + f(Pasture_{i.t}) + h(PriFor_{i.t}) + m(Temp_{i.t}) + n(Prec_{i.t}) + \varepsilon_{i.t}$	26.49	2.325
16	Env.Model.5	$V_{i.t} = a + b(Year_t) + d(Genotype_i) + e(Pop_{i.t}) + f(Pasture_{i.t}) + h(PriFor_{i.t}) + k(SecFor_{i.t}) + m(Temp_{i.t}) + n(Prec_{i.t}) + \varepsilon_{i.t}$	22.93	2.347
17	Env.Model.5.Red	$V_{i.t} = a + b(Year_t) + d(Genotype_i) + e(Pop_{i.t}) + f(Pasture_{i.t}) + m(Temp_{i.t}) + \varepsilon_{i.t}$	15.87	2.344

18	Env.Model.6	$V_{i.t} = a + b(Year_t) + \acute{d}(Region_i) + d(Genotype_i) + e(Pop_{i.t}) + f(Pasture_{i.t}) + h(PriFor_{i.t}) + k(SecFor_{i.t}) + m(Temp_{i.t}) + n(Prec_{i.t}) + \varepsilon_{i.t}$	32.09	2.313
19	Env.Model.6.Red	$V_{i.t} = a + b(Year_t) + \acute{d}(Region_i) + d(Genotype_i) + e(Pop_{i.t}) + f(Pasture_{i.t}) + h(PriFor_{i.t}) + m(Temp_{i.t}) + n(Prec_{i.t}) + \varepsilon_{i.t}$	30.20	2.312 *
20	Env.Model.7	$V_{i.t} = a + b(Year_t) + \acute{d}(Region_i) + d(Genotype_i) + e(Pop_{i.t} \times Region_i) + f(Pasture_{i.t} \times Region_i) + h(PriFor_{i.t} \times Region_i) + k(SecFor_{i.t} \times Region_i) + m(Temp_{i.t} \times Region_i) + n(Prec_{i.t} \times Region_i) + \varepsilon_{i.t}$	115.45	2.345
21	Env.Model.7.Red	$V_{i.t} = a + b(Year_t) + \acute{d}(Region_i) + d(Genotype_i) + e(Pop_{i.t}) + f(Pasture_{i.t}) + h(PriFor_{i.t} \times Region_i) + m(Temp_{i.t} \times Region_i) + \varepsilon_{i.t}$	62.55	2.311
22	Env.Model.8	$V_{i.t} = a + b(Year_t) + \acute{d}(Region_i) + d(Genotype_i) + e(Pop_{i.t} \times Genotype_i) + f(Pasture_{i.t} \times Genotype_i) + h(PriFor_{i.t} \times Genotype_i) + k(SecFor_{i.t} \times Genotype_i) + m(Temp_{i.t} \times Genotype_i) + n(Prec_{i.t} \times Genotype_i) + \varepsilon_{i.t}$	74.39	2.336
23	Env.Model.8.Red	$V_{i.t} = a + b(Year_t) + \acute{d}(Region_i) + d(Genotype_i) + e(Pop_{i.t} \times Genotype_i) + f(Pasture_{i.t}) + h(PriFor_{i.t} \times Genotype_i) + k(SecFor_{i.t}) + m(Temp_{i.t} \times Genotype_i) + n(Prec_{i.t} \times Genotype_i) + \varepsilon_{i.t}$	62.71	2.311
24	Env.Model.7.Red2	$V_{i,t} = a + b(Year_t) + \acute{d}(Region_i) + d(Genotype_i) + e\left(Pop_{i,t}\right) + f\left(Pasture_{i,t}\right) + m\left(Temp_{i,t} \times Region_i\right) + n\left(Prec_{i,t}\right) + \varepsilon_{i,t}$	44.22	2.318

References

1. Rodriguez, A. & Laio, A. Clustering by fast search and find of density peaks. Science 344, 1492-1496, doi:10.1126/science.1242072 (2014).
2. Perry, R. D. & Fetherston, J. D. Yersinia pestis-etiologic agent of plague. Clin Microbiol Rev 10, 35-66 (1997).
3. Pollitzer, R. & World Health Organization. Plague. (World Health Organization, 1954).
4. World Health Organization. Plague Manual: Epidemiology, Distribution, Surveillance and Control, 1999).