Supporting Information

Smith et al. 10.1073/pnas.0809145106

SI Text

Time Series Data for All Four Sites. See Fig. S1.

Simulation Tests of Parameter Estimation Methodology. We simulated Eq. **5**, 100 times, for all four sites over the same length of time as the time series. We used our data on S and R and the means of our best-fit parameters detailed in Fig. 3, as the equation parameters, reflecting the methodology used in the main study. Each simulation was initialized with the earliest time series estimates for the population density of infected individuals. The equation was integrated over monthly time steps by using the method outlined in the main article, and log normally distributed process error was added at the end of each 28-day period of integration. This modified value was used as the infected population density at the beginning of the next integration period, i.e., process error accumulates. Observation error was then applied to the whole time series: the susceptible and recovered population densities from the data, and our estimated population densities for infected individuals. In the absence of estimates of observation or process error, we used estimates of observation error from our capture-mark-recapture (CMR) estimation of total abundance. We assumed that observation error for the subcomponents was the same and estimated a log-normal observation error standard deviation of 0.13 $\log_e$ units. In the absence of other data, we assumed the same figure for the standard deviation of log normally distributed process error. We then applied our parameter estimation techniques to these simulated datasets.

Our analysis reveals that our methodology can accurately capture the original parameters used in the model despite significant process and observation error (Table S1). All 100 parameter estimation trials contained the true parameter within the 95% credibility intervals, for all five model parameters, and the deviation of the mean of these parameter estimates from the true mean was small (Table S1). Although these experiments do not exclude the possibility that temporally varying or non-log-normally distributed observation and process error may bias our parameter estimates, they do indicate prima facie evidence of robustness to substantial unmodeled process error.

Full Parameter Estimation Results. See main text for parameter definitions.

$$\frac{dI(t)}{dt} = \frac{\beta K^q S(t) I(t)}{N(t)^q} - M(1 + \alpha \sin(2\pi(t - \Delta)))I(t) \quad \textbf{[S1]}$$

This model combines death and recovery rate into a single combined seasonal function with a period of 1 year (Fig. S2).

$$\frac{dI(t)}{dt} = \frac{\beta K^q S(t) I(t)}{N(t)^q} - \gamma I(t) \quad \textbf{[S2]}$$

This model combines death and recovery rate into a constant rate parameter (Fig. S3).

$$\frac{dI(t)}{dt} = \frac{\beta K^q S(t) I(t)}{N(t)^q} - \frac{\beta S(\tau) I(\tau) \exp\left(-\int_\tau^t b(1 + \alpha \sin(2\pi(x - \sigma)))\partial x \right)}{N(\tau)^q} - b(1 + \alpha \sin(2\pi(t - \sigma)))I(t) \quad \textbf{[S3]}$$

This model assumes that recovery rate is exactly equal to infection rate 28 days previously, discounting for the number of infected individuals that died in the intervening time period. Death rate is assumed to be seasonal with a period of 1 year (Fig. S4).

$$\frac{dI(t)}{dt} = \frac{\beta K^q S(t) I(t)}{N(t)^q} - \frac{\beta S(\tau) I(\tau) \exp(-\lambda \varepsilon)}{N(\tau)^q} - \lambda I(t) \quad \textbf{[S4]}$$

This model assumes that recovery rate is exactly equal to infection rate 28 days previously, discounting for the number of infected individuals that died in the intervening time period. Death rate is assumed to be constant (Fig. S6).

Effects of Incorporating Seasonality into the Transmission Term of Eq. 5 on the Fit of the Model Predictions to the Data. Here we report more details on the relative improvement of the predictions of our best-fit model (Eq. **5** in the main text) by allowing the transmission coefficient β, or the density dependence of host contact rate q, to become seasonal functions of time. To allow for seasonal variation we converted these parameters into sinusoidally varying functions of time with a period of exactly one year and with 3 free parameters: the amplitude, the mean, and the phase shift. We estimated the most likely parameter values in the resulting model by using the same Markov Chain Monte Carlo (MCMC) model fitting procedure as described in the *Methods* in the main text. We report the change in mean $-\log_{10}$(likelihood) and the Deviance Information Criterion (DIC) relative to our previous best fit model (Eq. **5** in the main text).

Making β a function of time alone (q constant) reduced $-\log_{10}$(likelihood) by 4.1 units (95% C.I., −6.2 and −2.0) and DIC by 3.8 units. Making q a function of time alone (β constant) reduced $\log_{10}$(likelihood) by 7.8 units (95% C.I., −9.9 and −5.7) and DIC by 7.4 units. Making β and q functions of time reduced $-\log_{10}$(likelihood) by 9.9 units (95% C.I., −13 and −4.0) and DIC by 10 units (note that reductions in $-\log_{10}$(likelihood) correspond to an improvement in model fit to the data).

Theoretical Model of Field Vole–Cowpox Virus Interactions and the Effects of Adding Stochasticity. The theoretical model we studied to produce Fig. 2*B* is a straightforward modification of that investigated previously (1) to study the population dynamical consequences of cowpox virus–field vole interactions. The model is given by

$$\frac{dS(t)}{dt} = \Psi(t)(S(t) + fZ(t))(1 - k_1 N(t)) - \frac{\beta K^q S(t) I(t)}{N(t)^q} - \mu S(t), \quad \textbf{[S5a]}$$

$$\frac{dI(t)}{dt} = \frac{\beta K^q S(t) I(t)}{N(t)^q} - (\theta + \eta + \mu) I(t), \quad \textbf{[S5b]}$$

$$\frac{dY(t)}{dt} = \eta I(t) - (\omega - \mu) Y(t), \text{ and} \quad \textbf{[S5c]}$$

$$\frac{dZ(t)}{dt} = \omega Y(t) - \mu Z(t), \qquad \textbf{[S5d]}$$

with

$$\Psi(t) = \begin{Bmatrix} \psi & \mathrm{T} < t < \mathrm{T} + L \\ 0 & \mathrm{T} + L < t < \mathrm{T} + 1 \end{Bmatrix}, \text{ and}$$

$$N(t) = S(t) + I(t) + Y(t) + Z(t)$$

Here, L is the reproductive season length in units of a fraction of one year, T is time in integer years, and t is continuous time. The host population, $N(t)$, is divided into four classes of individual: those that are susceptible to infection $[S(t)]$, infected individuals that cannot reproduce $[I(t)]$, recovered and immune individuals that cannot yet reproduce $[Y(t)]$, and recovered and immune individuals that can reproduce $[Z(t)]$. We assume that disease free per capita death rate, μ, is constant throughout the year, but that the per capita birth rate is seasonal $[\Psi(t)]$ with no births possible in the nonreproductive season ($\Psi(t) = 0$) and a constant maximum per capita birth rate in the reproductive season ($\Psi(t) = \psi$). The birth rate is also assumed to be density-dependent and is modified because of a susceptibility to crowding coefficient, k_1. Infected individuals potentially have an increased mortality rate due to effects of the disease (θ), and recover at a constant rate (η). Recovered individuals initially enter an immune, but nonreproductive class that they leave at rate ω and regain a proportion of their reproductive capacity (f). Our modification to this set of equations (1) is that we have adopted our generalized transmission term (Eq. **4**), but we assume that k_2 is the equilibrium population size of hosts in an aseasonal environment (for all time). This means that in aseasonal environments (for all time) the equilibrium abundances of all population components are independent of q. The parameter values assumed to produce Fig. 1 are identical to those used in ref. 1 to produce their figure 1 (m–r) and are $f = 0.225$, $\psi = 10.628$, $\mu = 3.7$, $k_1 = 0.0026$, $L = 0.5833$, $\beta = 0.9$, $\theta = 4.3$, $\eta = 14$, $\omega = 1.884$, and $K = 38$.

This model was solved by using a standard differential equation solving algorithm (the MATLAB algorithm "*ode45,*" which is based on an explicit Runge-Kutta (*4,5*) formula) and the MATLAB code used is available from the corresponding author on request. In addition we have developed a software tool that can be used to solve these equations for user-defined sets of parameter values. This is downloadable from the Microsoft Research Cambridge Software and Tools for Computational Science web page (http://research.microsoft.com/en-us/groups/science/software.aspx) or is available from the author on request.

The analysis of this model presented in the main article was for demonstration purposes only. In Fig. S5 we give the results of the same simulations but with the birth rate parameter ψ perturbed by a normally distributed amount with a mean of zero and a standard deviation of 0.1, to illustrate the effects of stochasticity on our findings. This shows that, whereas much of the bifurcation structure is removed by the addition of noise, there are still notable changes in the character of the multiyear dynamics in terms of the dominant period, the mean population size at a given time of the year, and the amplitude of the multiyear variation.

Preliminary Analysis of the Performance of the Best-Fit Model over Longer-Term Simulations. Here we give the results of our preliminary analysis of the performance of our best-fit model over longer-term simulations, included for purposes of illustration. We performed the simulations in two different ways. In the first we followed the general approach of our other methods and assumed that the dynamics of susceptible and recovered individuals through time are exactly as determined by our data. We have taken this approach throughout, because we have no models of the dynamics of these subcomponents that have been tested with data. In the second approach we modified our theoretical model, detailed in the preceding section, to see whether the simulated dynamics show any similarities to the actual data. We present our data as qualitative illustrations and leave formal quantitative analysis and model comparisons to future studies.

For the first analysis we selected 100 parameter combinations at random from the MCMC model-fitting exercise (transients removed) and simulated the data in exactly the same way as detailed in *Simulation Tests of Parameter Estimation Methodology*, but with noise due to process and observation error removed. The results of these simulations for one site only are shown in Fig. S7*A* (we only present data for one site because of *SI Text* space constraints). We also repeated these simulations by using the parameters from the MCMC fitting exercises where we fit Eq. **5**, but assuming $q = 0$ and $q = 1$ (examples in Fig. S7 *B* and *C*). These simulations indicate that the model performs well at simulating the dynamics over a complete 6-year window, especially given that it has only previously been tested against its ability to predict 28 days into the future. However, it was notable from the simulations that the model performs less well at certain times in the time series and for certain sites (omitted because of space constraints). Determining the reasons for these discrepancies is a natural area for future work. Comparing the predictions of the different models, those models with $q = 0$ and $q = 1$ predict similar qualitative dynamics, but with notable quantitative differences. We found that their fits to the data were visibly worse than those of the best-fit model, but have not performed any formal comparison.

For the second analysis we again selected 100 parameter combinations at random from each MCMC model-fitting exercise (transients removed). However this time we simulated the dynamics by using Eq. **S5** above, but with the transmission and loss terms changed to be those of our best-fit model, or those of the best-fit models assuming $q = 0$ and $q = 1$. We changed the timing of onset of the nonreproductive season such that it started on November 1 and ended on March 31, and arbitrarily set $Y(0)$ equal to 10% of the initial population size of recovered individuals in the dataset and $Z(0)$ to the remaining 90%. Otherwise we modified none of the other parameters, and no forms of stochasticity were added to the model. The simulations predict similar dynamics for all four sites so we only illustrate their predictions for one site (Fig. S8). The results indicate that the simulations perform surprisingly well considering that the model has never been previously tested with field data. Similar conclusions to those from the previous analysis can be drawn regarding the absolute predictive ability of these simulations and the model comparisons: the model performs less well at certain times in the time series and for certain sites and the different models predict similar qualitative dynamics, but with notable quantitative differences. Again, although the fits to the data of the models with $q = 0$ and $q = 1$ are visibly worse than those of the best model.

1. Smith M, et al. (2008) Disease effects on reproduction can cause population cycles in seasonal environments. *J Anim Ecol* 77:378–389.

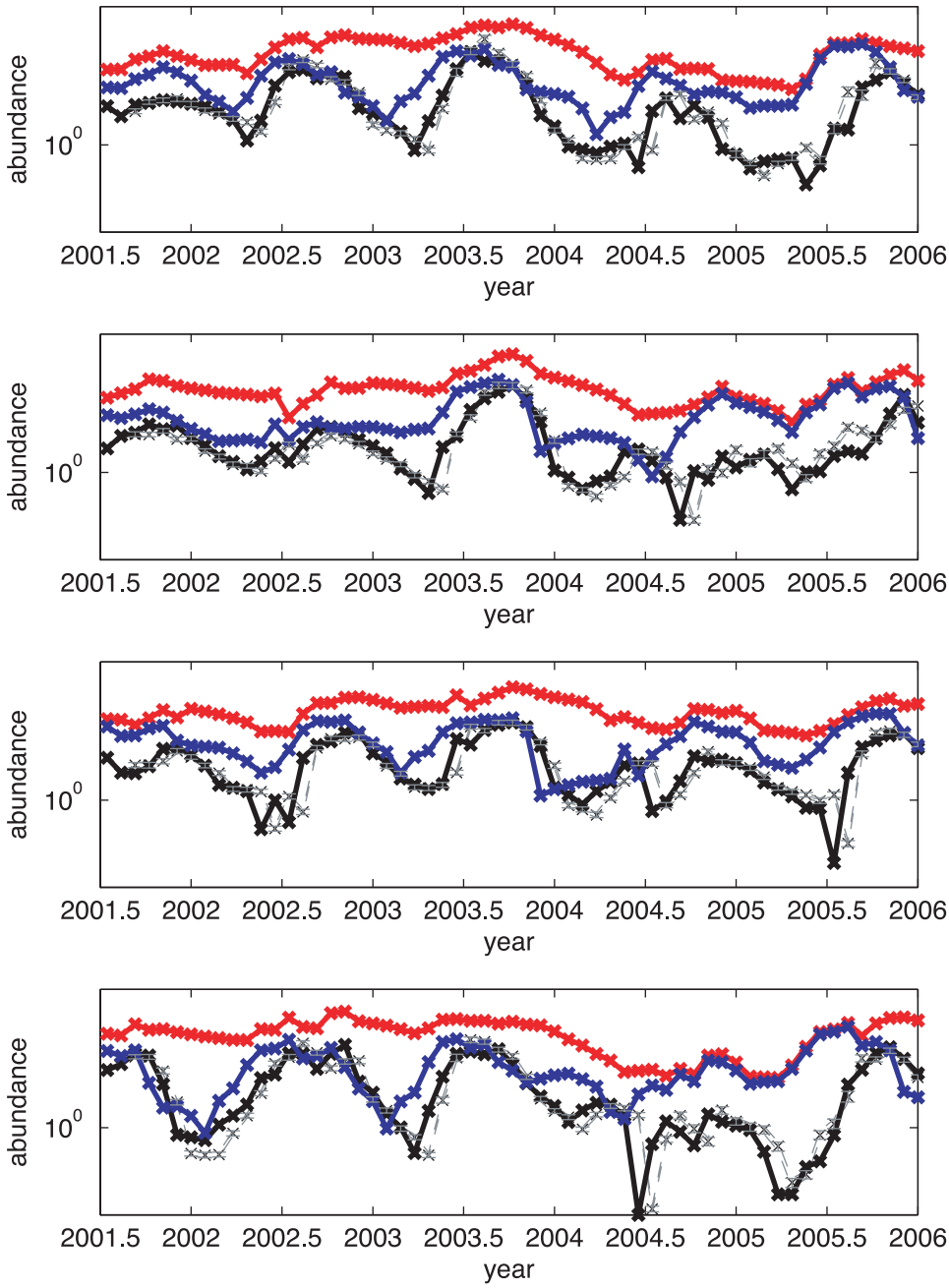

Fig. S1. Time series data of cowpox infection status in four natural populations of field voles. The format is identical to Fig. 1 of the main text except that all four populations are detailed. Refer to the legend of Fig. 1 for details.

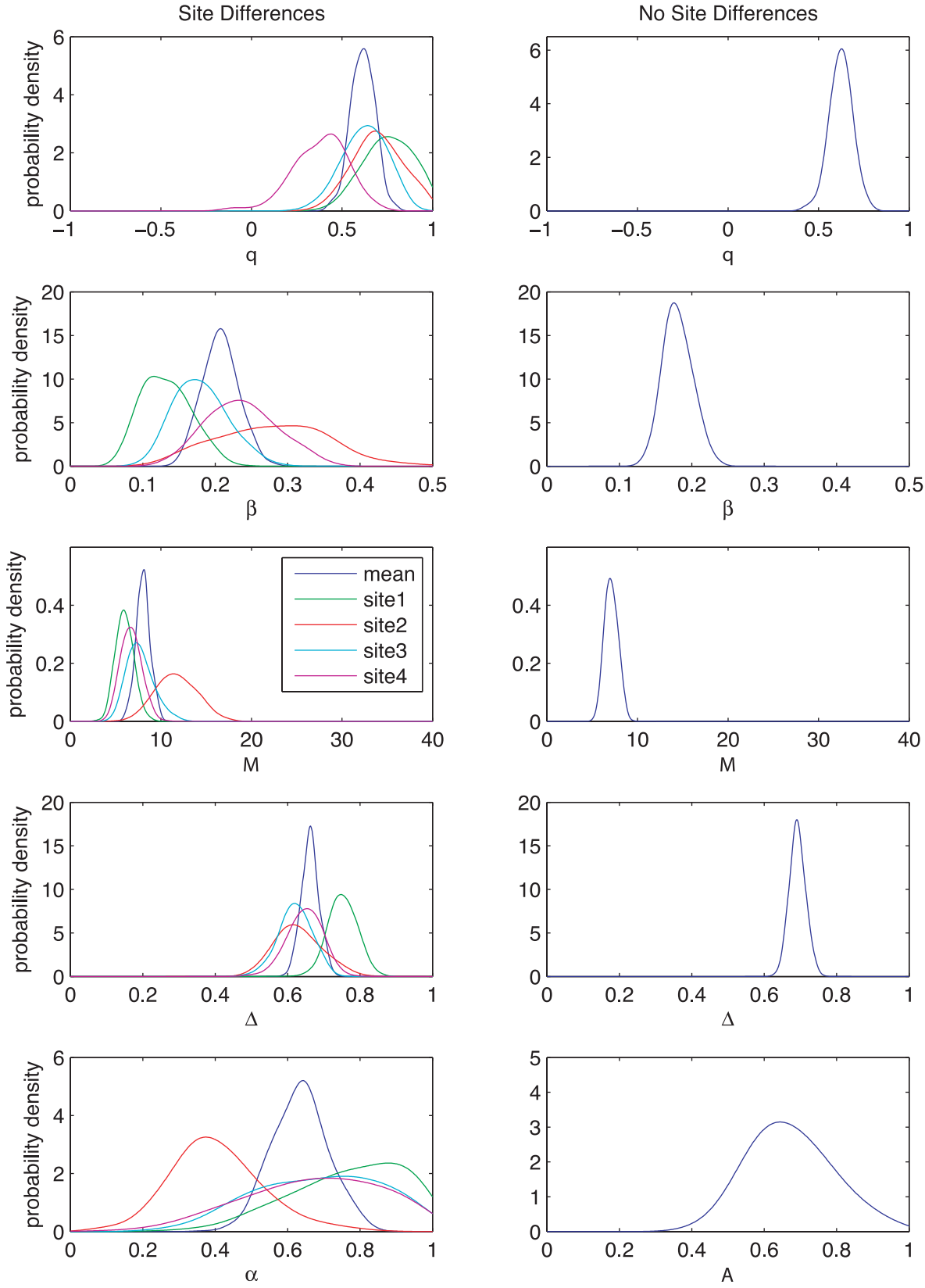

Fig. S2. Parameter estimates for Eq. **S1** from MCMC model fitting to time series data. See Table 1 legend (main text) for parameter definitions.

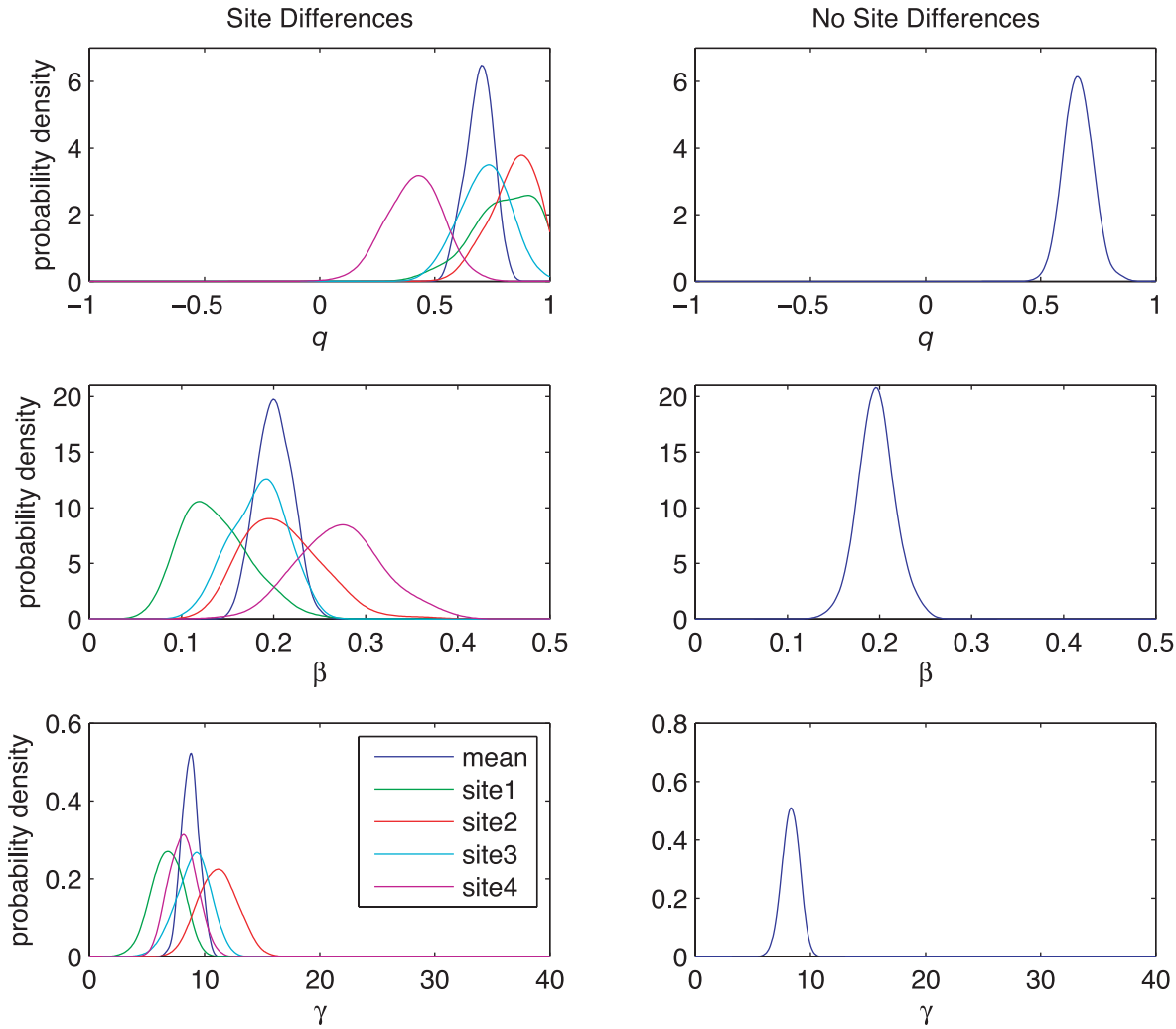

Fig. S3. Parameter estimates for Eq. **S2** from MCMC model fitting to time series data. See Table 1 legend (main text) for parameter definitions.

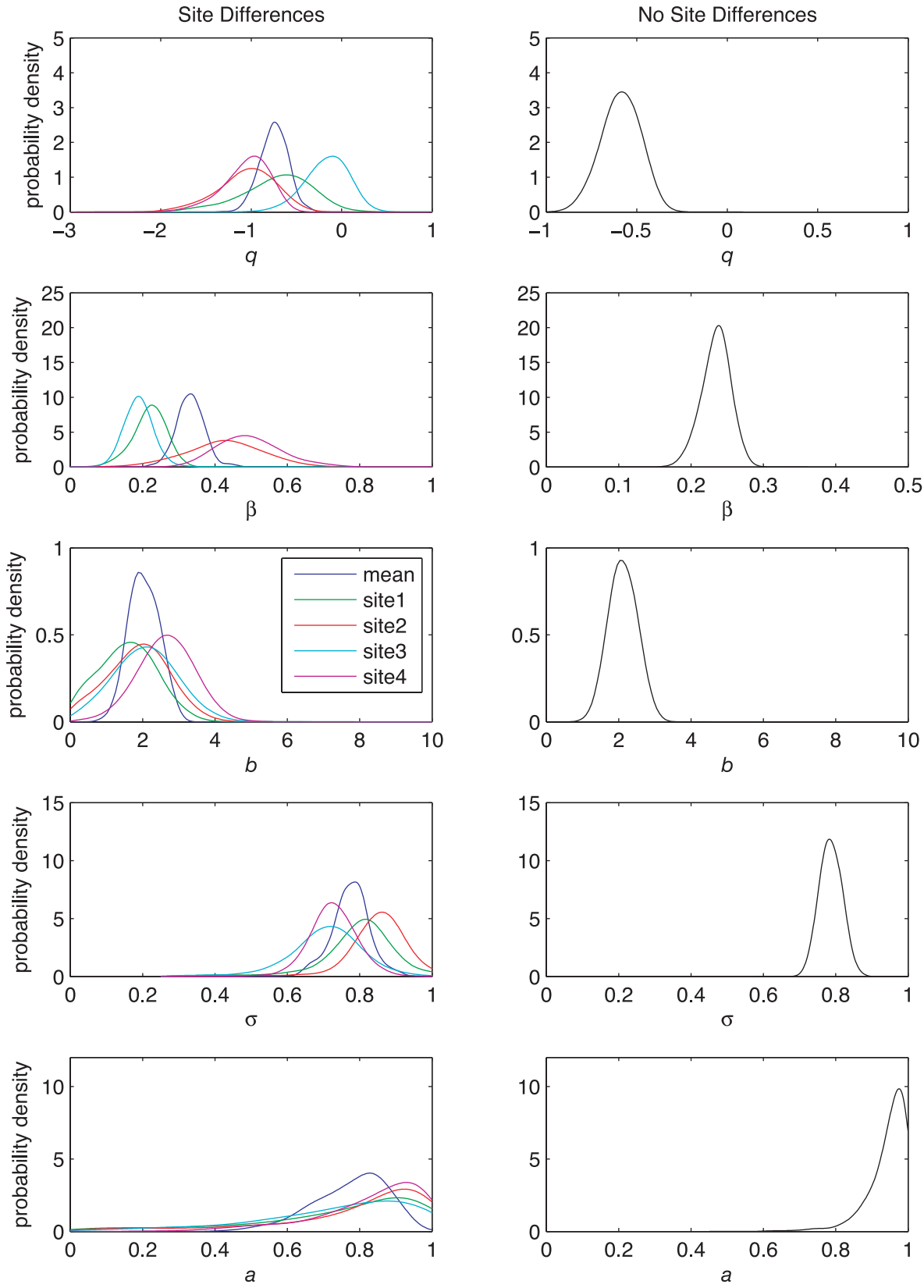

Fig. S4. Parameter estimates for Eq. **S3** from MCMC model fitting to time series data. See Table 1 legend for parameter definitions.

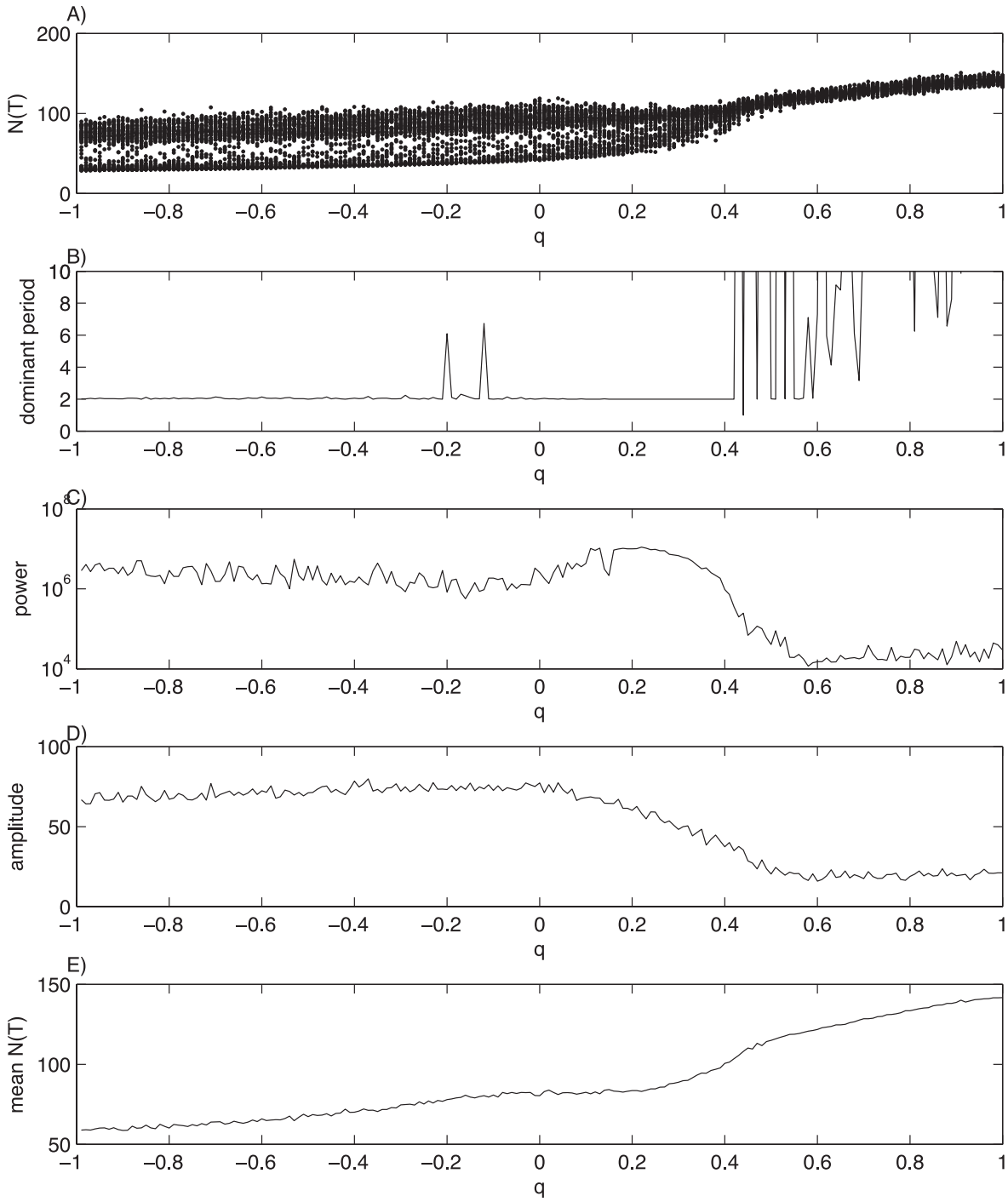

Fig. S5. The effects of nonlinear contact rate–abundance relationships in the presence of stochasticity. Predictions of the same model used to create Fig. 2 of the main text but with the birth rate parameter ψ perturbed by a normally distributed amount with a mean of zero and a standard deviation of 0.1. (*A*) The updated bifurcation diagram. $N(T)$ is the total population size at time T, the start of the reproductive season. (*B*) The dominant period of the data in the bifurcation diagram (256 years for each value of q). We used discrete Fourier transform to identify the dominant period, having set the maximum ''dominant period'' shown in the graph to be 10. For values of $q > 0.5$, the time series show quite complex dynamics and the strength of the dominant period is relatively weak compared with the other periods. This is illustrated in *C*, which shows the relative strengths of the dominant period (note the log axis). In *D*, we show the amplitude of the dominant period, simply determined as the difference between the maximum and minimum of the time series. In *E*, we show the average $N(T)$ in the time series over 256 years of simulated data.

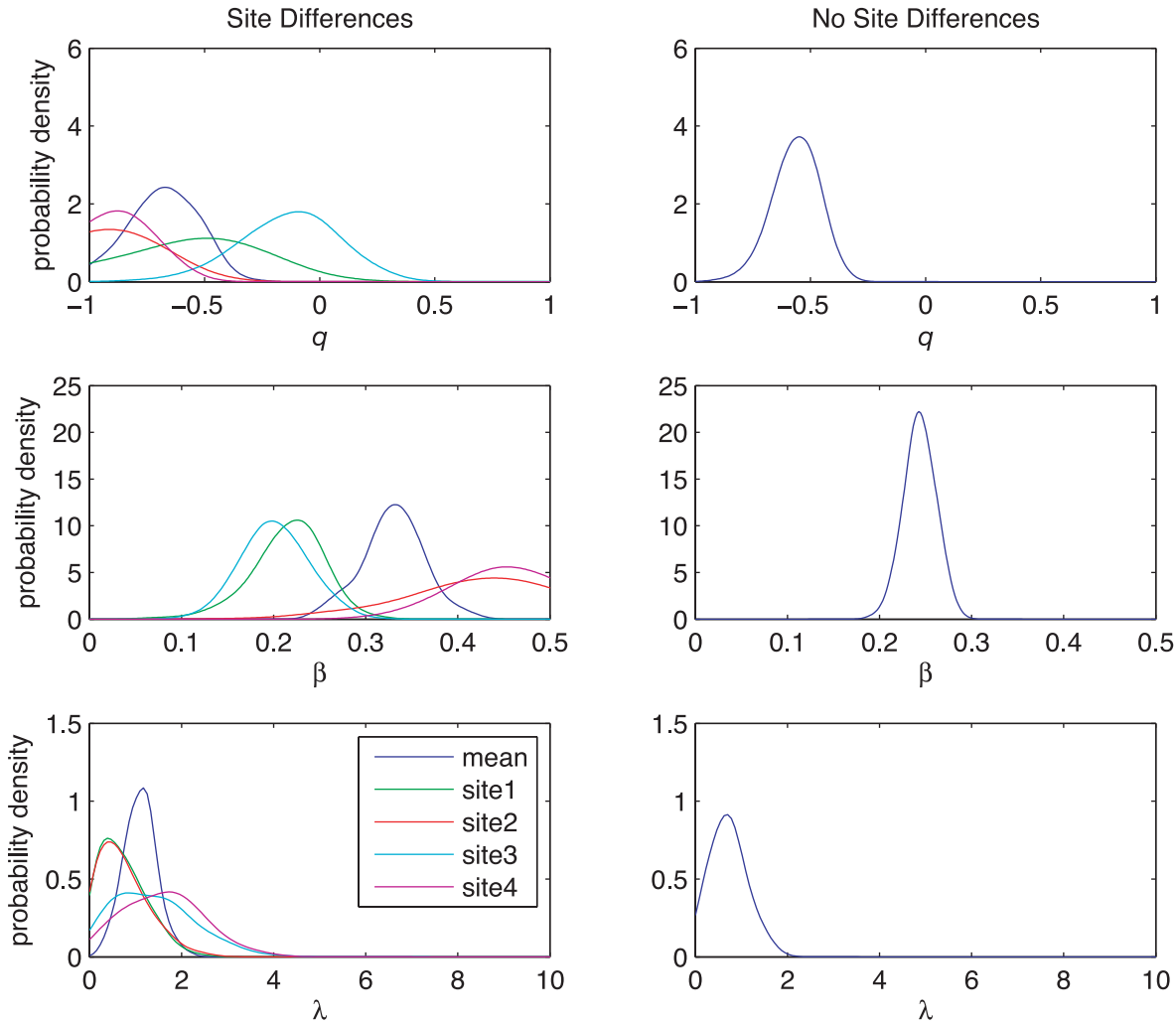

Fig. S6. Parameter estimates for Eq. **S4** from MCMC model fitting to time series data. See Table 1 legend (main text) for parameter definitions.

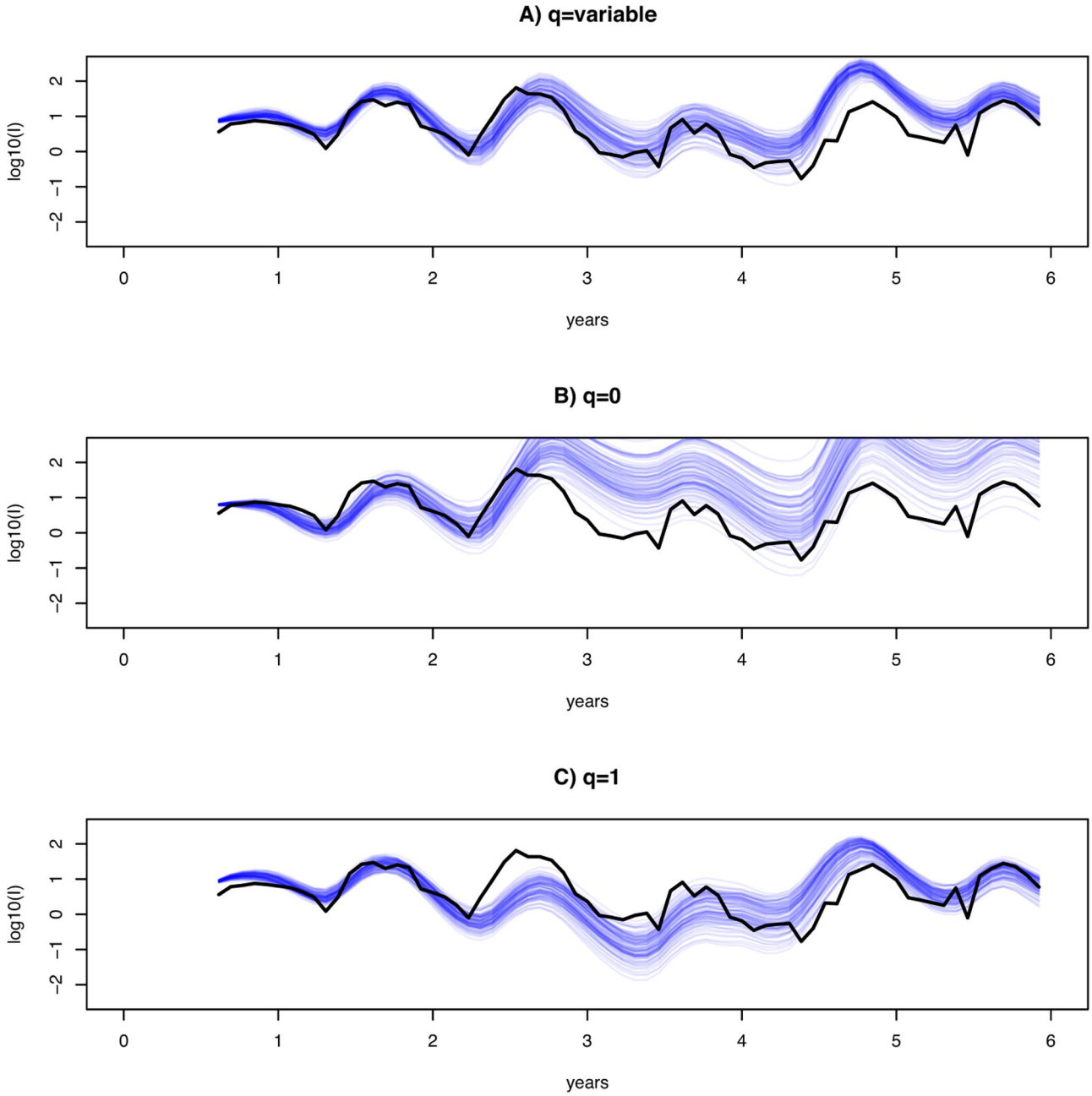

Fig. S7. Long-term predictions of the population density of field voles infected with Cowpox virus from our best-fit model, Eq. **5**, and from our models fit by MCMC while assuming $q = 0$ and $q = 1$. We used 100 randomly selected parameter combinations from our MCMC model-fitting exercise in combination with the time series of $S(t)$ and $R(t)$, and the initial value of infected individuals, $I(0)$. (*A*) Our best-fit model, Eq. **5**. (*B*) The best fit model with identical structure to Eq. 5 but with $q = 0$. (*C*) The best fit model with identical structure to Eq. **5** but with $q = 1$. Owing to space constraints, we show predictions for one site only.

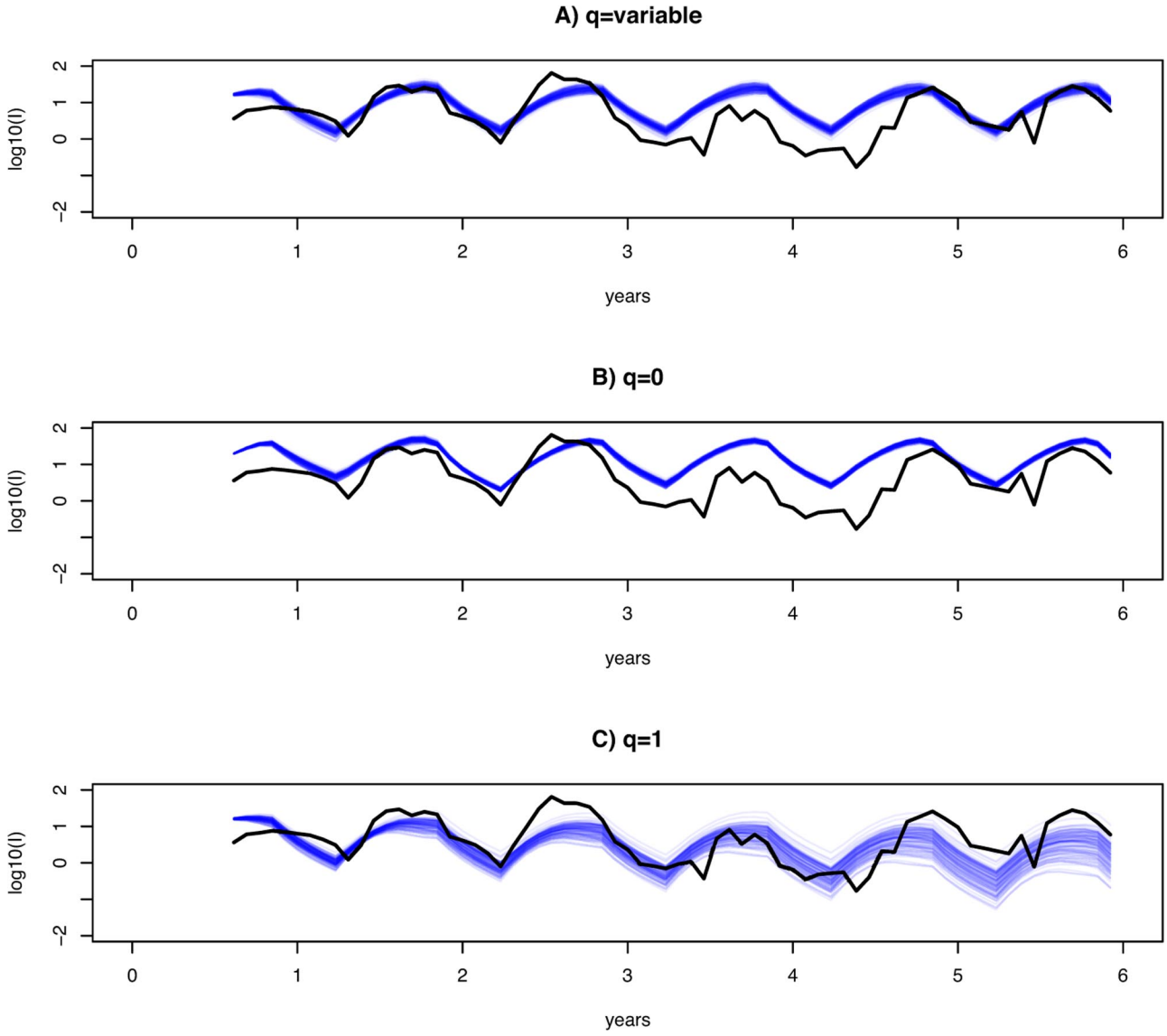

Fig. S8. Long-term predictions of the population density of field voles infected with Cowpox virus from the theoretical model. We modified the transmission and loss terms to be those of our best-fit model (Eq. **5**) (*A*) or those of the best-fit models where $q = 0$ (*B*) or $q = 1$ (*C*) is assumed. We used 100 randomly selected parameter combinations from our MCMC model fitting exercises in each case. We show predictions for one site only, because they are very similar for all four sites.

Table S1. Mean and root mean square deviation of parameter estimates from the true parameter values from our simulation experiments

Parameter	Mean deviation of estimated mean from true parameter	Root mean square deviation of estimated mean from true parameter
q	0.003	0.03
β	−0.002	0.01
Δ	0.001	0.008
M	0.02	0.05
α	−0.09	0.3

Parameters defined in Table 1 legend (main text).