SUPPLEMENTARY MATERIAL: LINKING SEX DIFFERENCES TO THE EVOLUTION OF INFECTIOUS DISEASE LIFE-HISTORIES

Matthew D. Hall[1] and Nicole Mideo[2]

[1] School of Biological Sciences and Centre for Geometric Biology, Monash University, Melbourne, VIC, 3800, Australia, matthew.hall@monash.edu

[2] Department of Ecology & Evolutionary Biology, University of Toronto, Toronto, ON, Canada, nicole.mideo@utoronto.ca

Model of between-host dynamics

We consider a simple model of the between-host dynamics of the *Daphnia-Pasteuria* system. This model (inspired by [46]) captures many of the features of the experimental system described in the main text, including decaying cadavers and the environmental reservoir of dormant, but infective spores. Using S, I, C and F to denote the densities of susceptible hosts, infected hosts, cadavers of infected hosts, and free-living spores of the pathogen, the epidemiological model we use is

$$\frac{dS(t)}{dt} = \theta - \mu S(t) - S(t)\sum_i \beta_i F_i(t) \quad \text{(S1a)}$$

$$\frac{\partial I_i(a,t)}{\partial t} = -\frac{\partial I_i(a,t)}{\partial a} - D_i(a) I_i(a,t) \quad \text{(S1b)}$$

$$\frac{dC_{i,a}(t)}{dt} = D_i(a) I_i(a,t) - \delta^C C_{i,a}(t) \quad \text{(S1c)}$$

$$\frac{dF_i(t)}{dt} = \int_0^\infty \rho \omega_i(a) C_{i,a} da - \delta_i^F F_i(t) \quad \text{(S1d)}$$

with boundary condition $I_i(0,t) = \beta_i F_i(t) S(t)$ and where i indexes pathogen genotype and a indexes the age of infection. We assume that there is a constant influx of susceptible individuals at rate θ and that all individuals suffer a constant per capita mortality rate of μ. The first term in equation S1b describes the flow of infected hosts who survive and move into the next infection age class. The term $D_i(a)$ is the infection-age and pathogen genotype-specific rate at which infections end due to host death; so, $D_i(a) = v_i(a) + \mu$, where $v_i(a)$ is the rate of pathogen-induced mortality at infection age a (*i.e.*, virulence). The cadavers of hosts that died at different infection ages are tracked separately; all cadavers decay at a rate δ^C (here, the superscript defines the class to which the death rate applies), but release spores at a rate that is proportional (with coefficient ρ) to the spore load at the time of host death, $\omega_i(a)$. The rates of environmental transmission and decay of spores are given by β_i and δ_i^F.

We can further simplify this model if we assume that the dynamics of decaying cadavers and spores in the environment are fast relative to the other epidemiological dynamics. When this is true, then $\frac{dC_{i,a}}{dt} = 0$ and $\frac{dF_i}{dt} = 0$, giving

$$C_{i,a}(t) = \frac{D_i(a)I_i(a,t)}{\delta^C} \tag{S2a}$$

$$F_i(t) = \frac{1}{\delta_i^F}\int_0^\infty \rho\omega_i(a)C_{i,a}(t)da. \tag{S2a}$$

We can then rewrite the model as:

$$\frac{dS(t)}{dt} = \theta - \mu S(t) - S(t)\sum_i \int_0^\infty \phi_i\omega_i(a)D_i(a)I_i(a,t)da \tag{S3a}$$

$$\frac{\partial I_i(a,t)}{\partial t} = -\frac{\partial I_i(a,t)}{\partial a} - D_i(a)I_i(a,t) \tag{S3b}$$

with boundary condition $I_i(0,t) = S(t)\int_0^\infty \phi_i\omega_i(a)D_i(a)I_i(a,t)da$, and where $\phi_i = \frac{\beta_i\rho}{\delta_i^F\delta^C}$. Although transmission and decay of spores may vary across pathogen genotypes in reality, for the remainder of our analyses we assume that they do not (thus, we drop the dependency of ϕ on i). From this general model, we now have a composite transmission function $B_i(a,t) = \phi\omega_i(a)D_i(a)S(t)$ which describes the per capita rate at which new infections of genotype i are created. We can then define the genotype specific transmission rate as $b_i(a) = \phi\ \omega_i(a)D_i(a)$. For linking this model to experimental data, we have direct estimates of $\omega_i(a)$; since we do not have independent experimental estimates of natural and pathogen-induced mortality, we equate our mortality estimates to virulence, with $D_i(a) \approx v_i(a)$. Finally, we arbitrarily assume that $\phi = 0.01$.

Evolutionary dynamics

With the between-host model in the form presented above, we can make use of the theoretical framework developed in Day et al [30] and refined in Mideo et al [25] to track the evolution of the mean transmission rate, $\bar{b}(a)$, and virulence, $\bar{v}(a)$, across pathogen genotypes in this system. Because the data we collected are discrete, we use the discrete-time equations for tracking evolutionary dynamics from [25].

Briefly, when pathogen fitness depends on two traits, like transmission and virulence (both functions of the age of infection), the evolutionary dynamics of those traits can be estimated as

$$\Delta\bar{b}(a) \approx \sum_{s=0}^{\infty}\psi_b(s;\bar{b},\bar{v})G_{b,b}(a,s) + \sum_{s=0}^{\infty}\psi_v(s;\bar{b},\bar{v})G_{b,v}(a,s) \tag{S4a}$$

$$\Delta\bar{v}(a) \approx \sum_{s=0}^{\infty}\psi_b(s;\bar{b},\bar{v})G_{v,b}(a,s) + \sum_{s=0}^{\infty}\psi_v(s;\bar{b},\bar{v})G_{v,v}(a,s). \tag{S4b}$$

In the above expressions, the $\psi_x(s;\bar{b},\bar{v})$ terms represent selection gradients acting on a particular trait (x) at a particular infection age (s), evaluated at the population mean trait values ($\bar{b},\bar{v}$). The $G_{x,y}(a,s)$ terms represent covariance matrices, describing any genetic correlation between traits x and y at infection ages a and s, respectively. These equations capture the fact that evolutionary change in a given trait at a particular age of infection (*e.g.*, $\bar{b}(a)$) is determined by the summed effect of selection acting on that trait at each age of infection (*e.g.*, $\psi_b(s;\bar{b},\bar{v})$), mediated by any genetic covariance in the trait across infection ages (*e.g.*, $G_{b,b}(a,s)$), plus any contribution of selection acting on the other trait at each infection age (*e.g.*,

$\psi_v(s;\bar{b},\bar{v})$), as mediated by any cross-covariance between traits (*e.g.*, $G_{b,v}(a,s)$). Note that we have dropped the dependencies of some of these terms on time, because we assume that the epidemiological dynamics are fast relative to the evolutionary dynamics (as in [25]). (Also note that we have used a different symbol (ψ) to specify selection gradients compared to [25] and [30].)

One novelty of the approach developed in [30] is that the selection gradient terms emerge from an epidemiological model. Using a model with a similar form to the one we presented above, these selection gradients were found to be

$$\psi_b(s;\bar{b},\bar{v}) = \frac{q(s)}{k}S \tag{S5a}$$

$$\psi_v(s;\bar{b},\bar{v}) = -\frac{q(s)}{k}\sigma(s), \tag{S5b}$$

where $q(s)$ gives the age distribution of infections (and is defined in the next section), k is a combined measure of the generation time of infections and transmission, and $\sigma(s)$ is the reproductive value of age infections [30]. The latter two quantities are given by

$$k = \frac{\sum_{a=1}^{\infty} a\bar{b}(a)S\lambda^{-a}\prod_{w=1}^{a}(1-\bar{v}(w))}{\sum_{a=1}^{\infty}\lambda^{-a}\prod_{w=1}^{a}(1-\bar{v}(w))} \tag{S6a}$$

$$\sigma(s) = \frac{\sum_{a=s}^{\infty}\bar{b}(a)S\lambda^{-a}\prod_{w=1}^{a}(1-\bar{v}(w))}{\lambda^{-s}\prod_{w=1}^{s}(1-\bar{v}(w))}, \tag{S6b}$$

where λ is the dominant eigenvalue of a matrix describing transitions between infections of different ages (the $\mathbf{L}$ matrix defined in the next section). Fuller derivations of these expressions can be found in [25] and [30]. Note that equations S5 are slightly different than the ones presented in [25] owing to the fact that mean virulence changes with infection age in our "Complex mortality" scenario (see dashed lines in Figure 3a, main text). Thus, the probability of an infection lasting to age a is given by $\prod_{w=1}^{a}(1-\bar{v}(w))$, rather than $(1-\bar{v})^a$ (as in [25]). When mean virulence is constant across infection ages, as in our "Simplified mortality" scenario, these terms are equivalent.

Substituting equations S5 and S6 into S4, gives System 1 of the main text.

Epidemiological settings

In general, we may expect quantities like S, $q(s)$, $\sigma(s)$ and k to change over time, but here (as in [25]) we assume that the epidemiological dynamics are fast relative to the evolutionary dynamics. To explore the influence of two different epidemiological settings, without having to explicitly track their dynamics, we define a matrix of the transitions between infections of different ages, similar to a Leslie matrix:

$$\mathbf{L} = \begin{pmatrix} S\bar{b}(1) & S\bar{b}(2) & \cdots & S\bar{b}(n-1) & S\bar{b}(n) \\ 1-\bar{v}(1) & 0 & \cdots & 0 & 0 \\ 0 & 1-\bar{v}(2) & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1-\bar{v}(n-1) & 0 \end{pmatrix}, \tag{S7}$$

where n is the maximum age of infection. Decomposition of the $\mathbf{L}$ matrix into its dominant eigenvalue and the associated eigenvector gives the long-term growth rate of infected hosts (λ) and the stable age distribution of infections ($q(s)$), respectively. Manipulation of λ, $q(s)$, and S values allows for the expanding epidemic and endemic scenarios to be estimated as in the main text.

As a final step, we calculate the rate of evolutionary change in our target trait of interest, $\bar{\omega}(a)$, the average spore loads at each age of infection in males and females. To a close approximation, for our data $\mathrm{E}[\omega(a)v(a)] = \mathrm{E}[\omega(a)]\mathrm{E}[v(a)]$. Thus, assuming $\bar{b}(a) = \phi\bar{\omega}(a)\bar{v}(a)$ (and $\frac{d\bar{\omega}(a)}{dt} \approx \Delta\bar{\omega}(a)$), we can estimate $\frac{d\bar{\omega}(a)}{dt}$ as follows:

$$\frac{d\bar{b}(a)}{dt} = \frac{d}{dt}\left(\phi\bar{\omega}(a)\bar{v}(a)\right) \tag{S8a}$$

$$\frac{d\bar{b}(a)}{dt} = \phi\left(\bar{v}(a)\frac{d\bar{\omega}(a)}{dt} + \bar{\omega}(a)\frac{d\bar{v}(a)}{dt}\right) \tag{S8b}$$

$$\frac{d\bar{\omega}(a)}{dt} = \frac{\left(\frac{1}{\phi}\frac{d\bar{b}(a)}{dt} - \bar{\omega}(a)\frac{d\bar{v}(a)}{dt}\right)}{\bar{v}(a)}. \tag{S8c}$$

The infection-age-specific average spore loads, $\bar{\omega}(a)$, and mortality rates, $\bar{v}(a)$, are given by the data. The other terms in equation S8c are approximated as solutions to system 1 in the main text.

Supplementary tables

Table S1: The Influence of age of infection on the production of transmission spores in male-limited and female-limited infections. Shown are the results of the regression analysis of spore loads (in millions per animal) on the age of infection (in days), partitioned by host sex and pathogen genotype. Spore loads at each time point were estimated by destructively sampled up to 20 infected individuals per sampling point.

Pathogen genotype	Intercept	SE	P-value	Slope	SE	P-value
Female-limited infections						
C24	-1.124	0.497	0.028	0.133	0.020	<0.001
C19	-0.208	0.209	0.321	0.106	0.008	<0.001
C20	0.302	0.327	0.358	0.082	0.011	<0.001
C14	1.221	0.363	0.001	0.074	0.011	<0.001
C18	0.526	0.331	0.115	0.097	0.009	<0.001
C01	-0.793	0.316	0.013	0.143	0.009	<0.001
Male-limited infections						
C24	-2.184	0.321	<0.001	0.129	0.015	<0.001
C19	0.357	0.120	0.003	0.029	0.005	<0.001
C20	0.246	0.123	0.047	0.031	0.005	<0.001
C14	0.426	0.135	0.002	0.021	0.005	<0.001
C18	0.198	0.148	0.181	0.027	0.005	<0.001
C01	0.307	0.112	0.007	0.025	0.004	<0.001

Table S2: The results of the two-factor analysis of covariance of the effect of host sex, pathogen genotype, age of infection and their interactions on the production of transmission spores. This analysis assessed whether the linear increase in spore loads per day was significantly different between the two sexes, the six genotypes, or all combinations. Before analysis, pathogen spore loads were square-root transformed and the analyses are based on a type 3 analysis of covariance.

Effect	df	F-ratio	P-value
Host sex (male, female)	1, 1554	63.481	<0.001
Pathogen genotype (G_P)	5, 1554	19.372	<0.001
Age of infection (in days)	1, 1554	162.488	<0.001
Sex × G_P	5, 1554	7.568	<0.001
Sex × Age	1, 1554	2.237	0.135
G_P × Age	5, 1554	14.467	<0.001
Sex × G_P × Age	5, 1554	7.360	<0.001

Table S3: Mortality estimates used in the main text. The last day that an infected individual could be tracked in the cross-infection experiment is given in brackets. We assumed a constant mortality rate from day 18 post-exposure (the first day that individuals were sampled) until this last day observed and determined the rate that would result in 99% of individuals dying by that day (rate =$\frac{\ln(0.01)}{last.day-17}$). Finally, we convert these estimates into a probability of death in a single time step (1-Exp[-rate]). Since we do not have independent estimates of virulence and natural mortality, we treat this composite measure as the trait that is evolving ($D_i(a) \approx v_i(a)$).

Scenario, Genotypes	Males	Females
Scenario 1 – simple mortality		
All Genotypes	0.3	0.2
Scenario 2 – complex mortality		
C24	0.264 (26)	0.401 (32)
C19	0.215 (32)	0.264 (36)
C20	0.181 (34)	0.237 (40)
C14	0.147 (36)	0.215 (46)
C18	0.138 (38)	0.197 (48)
C01	0.138 (38)	0.197 (48)