Theoretical Population Biology **62**, 199–213 (2002)
doi:10.1006/tpbi.2002.1595

TPB

The Evolution of Virulence in Vector-Borne and Directly Transmitted Parasites

Troy Day

Department of Zoology, University of Toronto, 25 Harbord St., Toronto, Ontario, Canada M5S 3G5
E-mail: dayt@zoo.utoronto.ca

Received November 2, 2001

Ewald (1994) has suggested that vector-borne parasites are expected to evolve a higher level of host exploitation than directly transmitted parasites, and this should thereby result in them being more virulent. Indeed, some data do conform to this general pattern. Nevertheless, his hypothesis has generated some debate about the extent to which it is valid. I explore this issue quantitatively within the framework of mathematical epidemiology. In particular, I present a dynamic optimization model for the evolution of parasite replication strategies that explicitly explores the validity of this hypothesis. A few different model assumptions are explored and it is found that Ewald's hypothesis has only qualified support as a general explanation for why vector-borne parasites are more virulent than those that are directly transmitted. I conclude by suggesting that an alternative explanation might lie in differences in inoculum size between these two types of transmission.

INTRODUCTION

Understanding why some parasites cause substantial host mortality while others cause very little has become one of the central issues in the application of evolutionary ideas to medicine (Williams and Nesse, 1991; Stearns, 1999; Trevathan *et al*., 1999). Such parasite-induced host mortality is typically taken as the definition of virulence in many studies (Bull, 1994; Levin, 1996; Frank, 1996; Ebert and Herre, 1996), and a number of hypotheses have been put forward to explain the evolution of different levels of virulence. One that has received a great deal of attention focuses on evolutionary trade-offs between fitness components of the parasite, and ignores evolutionary change in the host (Bull, 1994; Levin, 1996; Frank, 1996; Ebert, 1999). Much of this theoretical work is based on mathematical models that describe aspects of the ecological dynamics of the host–parasite interaction, but there have been some very influential and thought-provoking verbal models published as well.

Chief among these is an hypothesis put forward in a number of publication by Ewald (1983, 1991, 1994, 1995) stating that vector-borne (VB) parasites are expected to evolve a higher level of virulence than directly transmitted (DT) parasites. The reasoning is that, if high virulence (i.e., host mortality) is caused by high parasite replication within the host, then this high replication will also likely induce higher host illness, resulting in a lower host activity level. For vector-borne parasites (e.g., mosquito-borne parasites) this reduced activity level will likely have very little effect on the contact rate between hosts with respect to parasite transmission because it is a vector (e.g., a mosquito) that causes the contact between hosts. In other words, the contact rate between hosts from the perspective of parasite transmission is decoupled from host activity level per se. On the other hand, for directly transmitted parasites, this reduced activity level will translate directly into a reduced contact rate between hosts, thereby reducing the potential transmission of the parasite and imposing an additional cost to parasite

0040-5809/02 $35.00

replication that should select for the evolution of lower virulence.

Ewald has published various compilations of data for human parasites that provide empirical support for this prediction (1983, 1994), but interestingly, some skepticism has been voiced as to whether or not this hypothesis is logically sound (Van Baalen and Sabelis, 1995; Ewald, 1995; Read *et al*., 1999). The main issue raised is the question of whether or not this hypothesis would stand up to a more quantitative treatment. In particular, Ewald's analysis of this hypothesis has been largely verbal, whereas a large body of literature on virulence evolution is based explicitly on the mathematical models of epidemiology. It has been unclear whether such models would make predictions in accord with Ewald's hypothesis if they were extended to take account of differences in parasite transmission mode.

In (Day, 2001) I examined a related question by constructing some simple epidemiological models that allowed for a variety of transmission modes for horizontally transmitted parasites. That study identified two main selective factors that will likely affect the evolution of virulence in directly transmitted versus vector-borne parasites. The first is the additional cost of virulence to directly transmitted parasites that Ewald identified in his verbal models, and this selects for reduced virulence. The second is the effect of the timing of parasite transmission. During an infection the host activity level will drop, and this translates into a reduced transmission rate between hosts for DT parasites (but not for VB parasites). This curtails the effective lifespan of an infection in much the same way that an increase in disease-independent host mortality does (Anderson and May, 1982; Sasaki and Iwasa, 1991; Kakehashi and Yoshinaga, 1992; Lenski and May, 1994; Ebert and Weisser, 1997; Williams and Day, 2001), and this can select for higher replication and thereby higher virulence for DT parasites depending upon the way in which this drop in transmission rate occurs during the infection relative to the timing of parasite-induced mortality (Day, 2001). Consequently, this suggests that Ewald's hypothesis is valid only under certain assumptions.

One difficulty with the generality of this conclusion about Ewald's hypothesis, however, is that although the epidemiological framework presented in Day (2001) is quite general, the analysis of Ewald's hypothesis presented there made the simplifying assumption that parasite density within a host is constant during an infection. This assumption is implicit in many models of the evolution of virulence (for exceptions, see Diekmann *et al*., 1990; Sasaki and Iwasa, 1991; Anderson and May, 1991, Chapter 11; Antia *et al*., 1994; Levin *et al*., 1996), and it is made largely to simplify the analysis and predictions. From the perspective of Ewald's hypothesis, however, the timing of parasite replication and the concomitant effect on host mortality and contact rate are crucial in determining whether or not this hypothesis of valid. Therefore, a more complete and satisfying model should explicitly account for the parasite replication dynamics within a host, and the changes in host contact and mortality rate that this induces over an infection. Doing so necessitates using a more complex modeling approach based on dynamic optimization, and that is the primary goal of this article. Using such a model, I demonstrate that Ewald's hypothesis has only qualified validity as a general evolutionary explanation for why vector-borne parasites are more virulent than non-vector-borne parasites. I conclude by suggesting that, if VB parasites have larger inoculum sizes than DT parasites, then this might provide an alternative explanation for the differences in virulence that are observed.

GENERAL MODELING APPROACH

The model presented here is very similar to that of Sasaki and Iwasa (1991) and is based on the epidemiological framework presented in Day (2001). It begins with the fact that, in the absence of co- or superinfection (Bremermann and Pickering, 1983; Nowak and May, 1994; May and Nowak, 1995), the evolutionarily stable parasite strain is the one with the largest expected lifetime production of new infections generated by a single infected host, per susceptible host in the population (Diekmann *et al*., 1990; Frank, 1996). Denoting this by R, and allowing for the possibility that the transmission rate and parasite-induced mortality rate change during an infection, we have (Sasaki and Iwasa, 1991; Day, 2001)

$$R = \int_0^\infty \beta(N(t)) \exp\left\{ - \int_0^t (\delta + bN(s))\, ds \right\} dt, \quad (1)$$

where $\beta(N(t))$ and $bN(t)$ are the transmission rate and parasite-induced mortality rate at time t during the infection given the parasite has density $N(t)$ (where b is a constant), and δ is the (constant) disease independent host mortality rate. The form of equation (1) allows for the possibility that the parasite density within the host changes during an infection, thereby resulting in changes in transmission rate and parasite-induced mortality. Also note that Eq. (1) does not explicitly include the possibility that the host clears the infection through an

immune response, but if this clearance rate is constant, then we can subsume it in the parameter δ without any loss of generality.

In a great deal of the theory on virulence evolution, the instantaneous parasite-induced mortality rate is taken as the definition of virulence. I follow this approach in the present model, although this definition is slightly more problematic here because $bN(t)$ changes during an infection as parasite density changes. Nevertheless, if one parasite strain has a consistently higher value of $bN(t)$ than another throughout their respective infections (as is often the case here) we can conclude that it induces a higher level of mortality on the host (i.e., it has a higher virulence). It needs to be stressed, however, that because virulence is measured as *case mortality* in most of the data used to support Ewald's hypothesis, the use of instantaneous mortality rate as the definition of virulence in models aimed at explaining this data is appropriate, only if the rate of clearance of the disease through an immune response is the same for all parasite strains (Day, in press).

To formalize Ewald's hypothesis, I follow the approach of Day (2001) and decompose the transmission rate, β, into the product of two components: (1) the probability that a "contact" occurs between an infected individual and a susceptible individual, and (2) the probability of successful transmission given a contact occurs (Diekmann and Heesterbeek, 2000; Keeling and Grenfell, 2000). Here a 'contact' is any process that potentially takes parasites from an infected host to a susceptible host. This could be direct host–host contact but it could also be a vector (e.g., a mosquito) moving between hosts. Writing ϕ as the rate of contact occurrence and τ as the probability of successful transmission given a contact occurs, we then have $\beta = \phi\tau$.

Ewald's hypothesis rests on the notion that the rate of contact occurrence, as a function of parasite replication rate (and thereby virulence), is different depending upon the mode of transmission. Therefore, to model differences in transmission mode I simply specify ϕ as a function of the parasite replication strategy differently for VB versus DT parasites. In particular, I examine two different models, each of which makes this specification in a different way.

In the first model, I suppose that ϕ depends directly on parasite density within the host at each point during the infection; i.e., $\phi(N(t))$. In the second, I suppose that the dependence of ϕ on parasite density within the host occurs indirectly. More specifically, I suppose that ϕ changes through time, but its dynamics are governed by a differential equation that is a function of N. This second model is probably more realistic because the first supposes that ϕ is instantaneously tied to the parasite density, N (which is clearly not true). Moreover, it might be the case that ϕ sometimes decreases during an infection even if the parasite density within the host remains constant, and only the second model can incorporate this effect. Nevertheless, the first model is still useful because it is easier to analyze and understand, and it provides a benchmark against which to judge the results of the second model. Also, Day (2001) presented a simple example using these two types of dependencies under the assumption that parasite density within the host must be "chosen" at the beginning of an infection and that it remains constant for all time. Thus, using these same forms here allows one to see how incorporating the additional realism that parasite density within a host changes through time affects the validity of Ewald's hypothesis.

I assume that τ is a function of parasite density, with the restrictions that $\tau(0) = 0$, and τ is unimodal with an intermediate maximum at some (possibly very large) value of N. This assumption differs slightly from that of Day (2001) (which assumed that τ was always increasing but at a diminishing rate), but this change does not impose any significant biological restriction (and it does make some of the proofs in the appendix easier). Also notice that I assume that differences in transmission mode do not affect the function τ (i.e., transmission mode affects the contact rate, ϕ, only). This simplification can easily be relaxed in the results that follow, and I will point how this might be done within the context of each section below. Finally, I follow Sasaki and Iwasa (1991), and assume that each parasite is characterized by its rate of replication throughout the infection, $r(t)$, and that the dynamics are simply $dN/dt = r(t)N$. I seek the replication schedule, $r^*(t)$, that is an evolutionarily stable (ES) strategy (i.e., it maximizes R), and then determine the level of virulence to which this corresponds.

DIRECT DEPENDENCE OF ϕ ON N

When there is a direct dependence of ϕ on N, in general we have $\beta(N) = \tau(N)\phi(N)$. I assume that $\phi(N)$ is either strictly decreasing (i.e., higher parasite densities lead to lower contact rates) or it is independent of N (parasite density has no effect on contact rate). Consequently, since τ has an intermediate maximum, $\beta(N)$ is always bounded between 0 and $\beta_{\max} < \infty$, and it

reaches its maximal value for some finite value of N. For this model, the fitness expression that is to be maximized (i.e. [1]) can be written more explicitly as

$$R(N_0; r) = \int_0^\infty \beta(N(t)) \exp\left\{ - \int_0^t (\delta + bN(s))\, ds \right\} dt \quad (2)$$

subject to the differential equation governing the dynamics of parasite replication within the host;

$$\frac{dN}{dt} = r(t)N, \quad N(0) = N_0, \quad (3)$$

where I assume $0 \leqslant r(t) \leqslant r_{\max}$. The notation $R(N_0; r)$ reflects the fact that fitness will depend, not only on the parasite's replication strategy, r, but on the inoculum size, N_0, as well.

It turns out to be considerably simpler to work with an alternative functional that represents the reproductive value of the infection. In particular, consider some time, $\hat{t}$, at which the parasite density within the host is given by $\hat{N}$. The expected number of new infections produced in the future, given the infection has survived up until time $\hat{t}$ (i.e., the reproductive value) is given by

$$V(\hat{N}; r) = \frac{1}{l(\hat{t})} \int_{\hat{t}}^\infty \beta(N(t)) l(\hat{t}) \exp\left\{ - \int_{\hat{t}}^t (\delta + bN(s))\, ds \right\} dt \quad (4)$$

with

$$\frac{dN}{dt} = r(t)N, \quad N(\hat{t}) = \hat{N}, \quad (5)$$

where $l(t) = \exp\{ - \int_0^t (\delta + bN(s))\, ds \}$ is the probability that the infection survives until time t. Notice that, because $l(\hat{t})$ cancels out of Eq. (4), V does not depend explicitly on $\hat{t}$. Importantly, determining the strategy, r, that maximizes reproductive value (4) at each point in time is equivalent to determining the strategy that maximizes total reproductive output (2). I use this fact below and solve for $r^*(t)$ by analyzing the dynamic programming equation for V.

Defining

$$V^*(\hat{N}) = \max_{\{r(t): t \geqslant \hat{t}\}} [V(\hat{N}; r)], \quad (6)$$

we can derive the dynamic programming equation by dividing the time interval $[\hat{t}, \infty)$ in the two disjoint intervals $\hat{t}$ to $\hat{t} + \Delta t$ and $\hat{t} + \Delta t$ to ∞, where Δt is assumed to be small. Thus we can write (6) as

$$V^*(\hat{N}) = \max_{\{r(t): t \geqslant \hat{t}\}} \begin{bmatrix} \int_{\hat{t}}^{\hat{t}+\Delta t} \beta(N(t)) \exp\left\{ - \int_{\hat{t}}^t (\delta + bN(s))\, ds \right\} dt \\ + \dfrac{l(\hat{t} + \Delta t)}{l(\hat{t})} \dfrac{1}{l(\hat{t} + \Delta t)} \int_{\hat{t}+\Delta t}^\infty \beta(N(t)) l(\hat{t}) \\ \times \exp\left\{ - \int_{\hat{t}}^t (\delta + bN(s))\, ds \right\} dt \end{bmatrix} \quad (7)$$

$$= \max_{\{r(t): t \geqslant \hat{t}\}} [\beta(N(\hat{t}))\Delta t + \{1 - (\delta + b\hat{N})\Delta t\} \times V(N(\hat{t} + \Delta t); r) + o(\Delta t)] \quad (8)$$

$$= \max_{\{r(t): t \geqslant \hat{t}\}} \left[\beta(N(\hat{t}))\Delta t + V(\hat{N}; r) - (\delta + b\hat{N})\Delta t V(\hat{N}; r) + \frac{dV}{dN} r \Delta t \hat{N} + o(\Delta t) \right], \quad (9)$$

which, upon simplifying, dividing by Δt, and taking the limit as $\Delta t \to 0$, gives

$$0 = \max_{\{r(t): t \geqslant \hat{t}\}} \left[\beta(\hat{N}) - (\delta + b\hat{N}) V(\hat{N}; r) + \frac{dV}{dN} r \hat{N} \right], \quad (10)$$

subject to differential equation (5).

Finally, we also have the boundary conditions $V^*(0) = 0$ and $\lim_{N \to \infty} V^*(N) = 0$. The first condition follows from the fact that the reproductive value of an infection must be zero if there are no parasites present within the host. The second condition follows from the fact that, because transmission rate is assumed to have some maximal value at an intermediate parasite density, the reproductive value of an infection must eventually decline to zero as $N \to \infty$ since host mortality rate becomes infinite in this case.

CHARACTERIZING THE EVOLUTIONARILY STABLE REPLICATION STRATEGY

To characterize the evolutionarily stable replication strategy, it is useful to first define $W(N)$ and N^* as follows:

$$W(N) \equiv \frac{\beta(N)}{\delta + bN}, \quad (11)$$

and N^* is the value of N that satisfies the equation

$$\frac{dW}{dN} = 0. \quad (12)$$

With the restrictions on β mentioned earlier, N^* represents the (unique) value of N that maximizes the

function W; i.e., there is only one value of N^* that satisfies Eq. (12) and this occurs where W is maximized. Notice that $W(N)$ is equal to the reproductive value of an infection with parasite density N, given that this density remains constant at N for all time.

The results of Appendix A demonstrate that the ES replication strategy satisfies the following conditions:

$$\text{if } N<N^* \quad \text{then } r^* = r_{\max}, \tag{13}$$

$$\text{if } N = N^* \quad \text{then } r^* = 0, \tag{14}$$

$$\text{if } N > N^* \quad \text{then } r^* = 0. \tag{15}$$

Conditions (13)–(15) specify the optimal replication strategy as a function of parasite density within the host. If $N_0<N^*$, then the parasite will first replicate a maximal speed until reaching density N^*, at which point it will cease replicating (see also Sasaki and Iwasa, 1991). These results can now be used to test the validity of Ewald's hypothesis in the context of this very simple model.

Testing the Validity of Ewald's Hypothesis

To test the validity of Ewald's hypothesis we need to specify functions for the contact rate, ϕ, for VB and DT parasites. From Ewald's verbal reasoning, I suppose that ϕ decreases with increased parasite density for DT parasites, but that it remains constant for VB parasites. This choice reflects that fact that we might expect the contact rate to decrease with increased parasite density for DT parasites because host individuals might become less active if they have higher parasites burdens. On the other hand, for VB parasites, because a vector provides the transmission, even if a host becomes less active the contact rate might nevertheless remain the same. With these choices we have

$$W_{\mathrm{DT}}(N) \equiv \frac{\tau(N)}{\delta + bN}\,\phi(N), \tag{16}$$

$$W_{\mathrm{VB}}(N) \equiv \frac{\tau(N)}{\delta + bN}\,\phi_0, \tag{17}$$

where $\phi(N)$ is a decreasing function of N, and ϕ_0 is a constant.

From these specifications we can see that

$$W_{\mathrm{DT}}(N) = \frac{\phi(N)}{\phi_0}\,W_{\mathrm{VB}}(N), \tag{18}$$

and because $\phi(N)$ is decreasing in N, we have $N^*_{\mathrm{DT}}<N^*_{\mathrm{VB}}$. Therefore, the density at which the parasite should stop replicating is lower for DT parasites than for VB parasites. This implies that, all else equal, VB parasites should replicate longer and reach higher densities than DT parasites, thereby resulting in a larger parasite-induced mortality rate. In other words, given this simple model, Ewald's hypothesis *is* valid.

To close this section, I note how this result changes if τ depends on transmission mode. Even more generally, if the probability of transmission given a contact occurs, τ, changes with transmission mode, and if the contact rate ϕ is a function of within-host parasite density for both transmission modes, then Eq. (18) becomes

$$W_{\mathrm{DT}}(N) = \frac{\tau_{\mathrm{DT}}(N)\phi_{\mathrm{DT}}(N)}{\tau_{\mathrm{VB}}(N)\phi_{\mathrm{VB}}(N)}\,W_{\mathrm{VB}}(N). \tag{19}$$

In this case, the model predicts that VB parasites will still be more virulent than DT parasites provided that

$$\frac{d}{dN}\left[\frac{\tau_{\mathrm{DT}}(N)\phi_{\mathrm{DT}}(N)}{\tau_{\mathrm{VB}}(N)\phi_{\mathrm{VB}}(N)}\right]_{N=N^*_{\mathrm{VB}}}<0. \tag{20}$$

Indirect Dependence of ϕ on N

To specify an indirect dependence of ϕ on N, I now suppose that ϕ changes over the course of an infection, and that its dynamics are governed by a differential equation that, itself, depends upon N. In particular, for simplicity I suppose that ϕ is non-increasing over the infection, starting at an initial contact rate, ϕ_0, and eventually dropping to a lower contact rate, k, that satisfies $0\leqslant k\leqslant\phi_0$ (provided the infection lasts long enough). Moreover, I assume that the rate of decrease in ϕ depends on the parasite density within the host. In general, one might want to allow the contact rate to increase at times during the infection, but given that parasite density is non-decreasing in the above model, this restriction seems reasonable. Nevertheless, it would be worthwhile to explore other possibilities, including an explicit account of immune system dynamics in future work.

Mathematically, I model the above assumptions by assuming that ϕ satisfies the differential equation

$$\frac{d\phi}{dt} = -\alpha(N)(\phi - k), \quad \phi(0) = \phi_0, \tag{21}$$

where $\alpha(N)$ is a non-decreasing function of N; higher parasite burdens typically result in a faster drop in the contact rate. Notice that this formulation allows for the possibility that ϕ decreases through time even if the parasite density is constant. For example, this might occur if the host waited some period of time before eventually "giving in" to the illness and reducing its activity level.

With these specifications we have $\beta(N) = \tau(N)\phi(t)$, and Eq. (1) is

$$R(N_0, \phi_0; r) = \int_0^\infty \tau(N(t))\phi(t) \times \exp\left\{-\int_0^t (\delta + bN(s))\, ds\right\} dt \quad (22)$$

subject to the differential equation

$$\frac{dN}{dt} = r(t)N, \quad N(0) = N_0, \quad (23)$$

where $0 \leqslant r(t) \leqslant r_{\max}$, and subject to differential equation (21).

As in the previous section, consider some time, $\hat{t}$, at which the parasite density within the host is given by $\hat{N}$, and the contact rate is given by $\hat{\phi}$. The expected number of new infections produced in the future, given the infection has survived up until time $\hat{t}$ (i.e., the reproductive value) is given by

$$V(\hat{N}, \hat{\phi}; r) = \frac{1}{l(\hat{t})} \int_{\hat{t}}^\infty \tau(N(t))\phi l(\hat{t}) \times \exp\left\{-\int_{\hat{t}}^t (\delta + bN(s))\, ds\right\} dt \quad (24)$$

with

$$\frac{dN}{dt} = r(t)N, \quad N(\hat{t}) = \hat{N}, \quad (25)$$

$$\frac{d\phi}{dt} = -\alpha(N)(\phi - k), \quad \phi(\hat{t}) = \hat{\phi}. \quad (26)$$

Notice that, again because $l(\hat{t})$ cancels out of Eq. (24), V does not depend explicitly on $\hat{t}$.

We can then derive the dynamic programming equation as done previously:

$$V^*(\hat{N}, \hat{\phi}) = \max_{\{r(t):t \geqslant \hat{t}\}} \left[\begin{array}{l} \int_{\hat{t}}^{\hat{t}+\Delta t} \tau(N(t))\phi \exp\left\{-\int_{\hat{t}}^t (\delta + bN(s))\, ds\right\} dt \\ \quad + \dfrac{l(\hat{t}+\Delta t)}{l(\hat{t})} \dfrac{1}{l(\hat{t}+\Delta t)} \int_{\hat{t}+\Delta t}^\infty \tau(N(t))\phi l(\hat{t}) \\ \quad \times \exp\left\{-\int_{\hat{t}}^t (\delta + bN(s))\, ds\right\} dt \end{array}\right]$$

$$\approx \max_{\{r(t):t \geqslant \hat{t}\}} [\tau(N(\hat{t}))\hat{\phi}\Delta t + \{1 - (\delta + b\hat{N})\Delta t\} V(N(\hat{t} + \Delta t), \phi(\hat{t} + \Delta t); r) + o(\Delta t)]$$

$$= \max_{\{r(t):t \geqslant \hat{t}\}} \left[\begin{array}{l} \tau(N(\hat{t}))\hat{\phi}\Delta t + V(\hat{N}, \hat{\phi}; r) \\ \quad -(\delta + b\hat{N})\Delta t V(\hat{N}, \hat{\phi}; r) + \dfrac{\partial V}{\partial N} r\Delta t\hat{N} \\ \quad -\dfrac{\partial V}{\partial \phi}\alpha(\hat{N})(\hat{\phi} - k)\Delta t + o(\Delta t) \end{array}\right],$$

which, upon simplifying, dividing by Δt, and taking the limit as $\Delta t \to 0$ gives

$$0 = \max_{\{r(t):t \geqslant \hat{t}\}} \left[\tau(\hat{N})\hat{\phi} - (\delta + b\hat{N})V(\hat{N}, \hat{\phi}; r) + \frac{\partial V}{\partial N} r\hat{N} - \frac{\partial V}{\partial \phi}\alpha(\hat{N})(\hat{\phi} - k)\right], \quad (27)$$

subject to differential Eqs. (25) and (26). Note that V^* must also satisfy the boundary conditions $V^*(0, \phi) \equiv 0$ and $\lim_{N\to\infty} V^*(N, \phi) = 0$.

CHARACTERIZING THE EVOLUTIONARILY STABLE REPLICATION STRATEGY

To characterize the evolutionarily stable replication strategy, I first define $W(N, \phi)$ and $N^*(\phi)$ as follows:

$$W(N, \phi) \equiv \frac{\tau(N)\phi}{\mu(N)} \frac{(k/\phi)\alpha(N) + \mu(N)}{\alpha(N) + \mu(N)}, \quad (28)$$

where $\mu(N) = \delta + bN$, and $N^*(\phi)$ is the value of N that, for that value of ϕ, satisfies

$$\frac{\partial W(N, \phi)}{\partial N} = 0. \quad (29)$$

Again, with biologically reasonable choices of α, $N^*(\phi)$ is the (unique) value of N that maximizes W as a function of the contact rate, ϕ; i.e., there is usually only one value of N that satisfies Eq. (12) for each value of ϕ, and this occurs where W is maximized (with respect to N). Again, $W(N, \phi)$ represents the reproductive value of an infection with density N (and contact rate ϕ) given that this density remains constant for all time.

Determining the ES replication strategy is more difficult in this model because of the additional state variable, ϕ. To get a qualitative understanding of the ES replication strategy, it is helpful to look at the $N - \phi$ phase plane. In fact, the qualitative nature of the ES strategy can be quite complicated since it depends on the form of the relationship between N^* and ϕ. To simplify matters, I will assume that $N^*(\phi)$ is either increasing for all ϕ or else it is decreasing for all ϕ. This is quite often the case for sensible choices of α, and the former case occurs when $\mu(N)/\alpha(N)$ increases with N while the latter occurs when $\mu(N)/\alpha(N)$ decreases with N (Appendix B). I characterize the ES replication strategy for each of these cases separately.

Consider the case where N^* increases with ϕ (i.e., $\mu(N)/\alpha(N)$ increases with N) and focus on a point $\{N, \phi\}$ in the $N - \phi$ plane (Fig. 1). The ES replication strategy

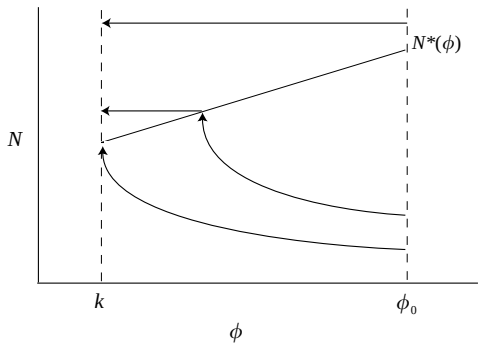

FIG. 1. A plot of the $N-\phi$ plane where $N^*(\phi)$ is increasing. The thin solid line represents the function $N^*(\phi)$ (it is shown as a straight line for simplicity, but it general it will be some curve). Thick solid arrows represent different trajectories in the $N-\phi$ plane under the optimal replication strategy. Points $\{N,\phi\}$ below the $N^*(\phi)$ curve result in $r^*=r_{\max}$ and thus N increases exponentially. Points $\{N,\phi\}$ above the $N^*(\phi)$ curve result in $r^*=0$ and therefore N is constant while ϕ decays to k.

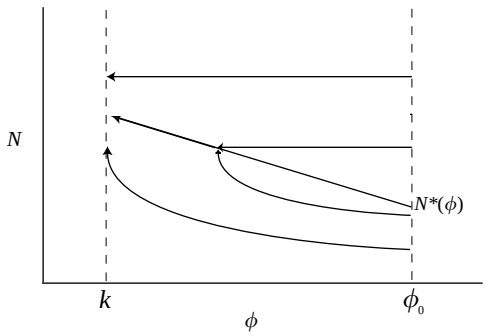

FIG. 2. A plot of the $N-\phi$ plane where $N^*(\phi)$ is decreasing. The thin solid line represents the function $N^*(\phi)$ (it is shown as a straight line for simplicity, but it general it will be some increasing curve). Thick solid arrows represent different trajectories in the $N-\phi$ plane under the optimal replication strategy. Points $\{N,\phi\}$ below the $N^*(\phi)$ curve result in $r^*=r_{\max}$ and thus N increases exponentially. Points $\{N,\phi\}$ above the $N^*(\phi)$ curve result in $r^*=0$ and therefore N is constant while ϕ decays. Points on the function $N^*(\phi)$ result in a singular control in which the trajectory exactly tracks this curve.

must satisfy the following conditions (Appendix C):

$$\text{if } N<N^*(\phi) \quad \text{then } r^*=r_{\max}, \tag{30}$$

$$\text{if } N>N^*(\phi) \quad \text{then } r^*=0. \tag{31}$$

The situation in which $N=N^*(\phi)$ occurs only at a point in time at the transition between (30) and (31).

Conditions (30) and (31) specify the ES replication strategy as a function of parasite density within the host and the contact rate. In particular, if condition (30) pertains, so that the point $\{N,\phi\}$ lies below the curve defined by $N^*(\phi)$ in the $N-\phi$ plane, then the parasite will first replicate at maximal speed (Fig. 1). Eventually, the trajectory intersects the $N^*(\phi)$ curve because the parasite density, N, reaches any finite density in finite time when $r^*=r_{\max}$ whereas the contact rate is non-increasing and bounded below by k. Once this occurs, the parasite ceases replication for all future time. Notice that, provided α is not equal to zero, the contact rate during this period continues to decrease until $\phi\to k$ (which happens asymptotically). On the other hand, if the initial point $\{N,\phi\}$ lies above the $N^*(\phi)$ curve, then the parasite will not replicate for all time (Fig. 1).

Now consider the case where N^* decreases with ϕ (i.e., $\mu(N)/\alpha(N)$ decreases with N) and focus on a point $\{N,\phi\}$ in the $N-\phi$ plane (Fig. 2). The ES replication strategy must satisfy the following condition (Appendix C):

$$\text{if } N<N^*(\phi) \quad \text{then } r^*=r_{\max}, \tag{32}$$

$$\text{if } N=N^*(\phi)$$
$$\text{then } r^*=\frac{\partial^2 W}{\partial\phi\partial N}\alpha(N)(\phi-k)\Big/\frac{\partial^2 W}{\partial N^2}N, \tag{33}$$

$$\text{if } N>N^*(\phi) \quad \text{then } r^*=0. \tag{34}$$

Clearly, this case is slightly more complex. If condition (32) pertains, so that the point $\{N,\phi\}$ lies below the curve $N^*(\phi)$, then again the parasite will replicate at maximal speed until the trajectory reaches this curve in the $N-\phi$ plane (Fig. 2). At this point condition (33) will pertain, and the ES replication strategy involves a so-called "singular control" during which the trajectory in the $N-\phi$ plane exactly follows that $N^*(\phi)$ curve to the point $N^*(k)$ (Appendix C). This replication strategy is specified mathematically in condition (33) in conjunction with Eqs. (27), (23) and (21). On the other hand, if the initial point $\{N,\phi\}$ lies above the $N^*(\phi)$ then there are two possibilities. If $N\geqslant N^*(k)$ (Fig. 2), then the parasite will not replicate ever (and the contact rate decays to k). If $N<N^*(k)$, then the parasite does not replicate until the trajectory in the $N-\phi$ plane intersects the $N^*(\phi)$ curve. At this point, condition (33) again pertains, and the parasite will employ a singular control, exactly tracking the $N^*(\phi)$ curve until $\phi\to k$ asymptotically (Appendix C; Fig. 2).

Testing the Validity of Ewald's Hypothesis

To test the validity of Ewald's hypothesis with this model, I specify the equation governing the time dynamics of ϕ (i.e., (21)) differently for VB and DT parasites. Again, from Ewald's verbal reasoning, I suppose that ϕ decreases with time according to Eq. (21) with some $\alpha(N)$ for DT parasites, whereas ϕ remains constant for VB parasites. The latter case

requires that $\alpha \equiv 0$. Again, this choice is meant to reflect that fact that we might expect the contact rate to decrease during an infection for DT parasites, and for it to do so at a faster rate when the parasite density is higher. On the other hand, for VB parasites the contact rate might remain roughly constant.

With these choices we have

$$W_{\mathrm{DT}}(N,\phi) \equiv \frac{\tau(N)\phi}{\mu(N)} \frac{(k/\phi)\alpha(N)+\mu(N)}{\alpha(N)+\mu(N)}, \qquad (35)$$

$$W_{\mathrm{VB}}(N,\phi) \equiv \frac{\tau(N)\phi_0}{\mu(N)}, \qquad (36)$$

where $\alpha(N)$ is an increasing function of N and ϕ_0 is the initial contact rate. Notice that $N^*_{\mathrm{DT}}(\phi)$ can be increasing or decreasing with ϕ depending on the choice of $\alpha(N)$, whereas $N^*_{\mathrm{VB}}(\phi)$ is independent of ϕ and equal to $N^*_{\mathrm{DT}}(k)$ (Fig. 3). Also notice that the depiction of $N^*_{\mathrm{VB}}(\phi)$ in the $N-\phi$ phase space is slightly misleading since it is assumed that ϕ never changes during an infection for VB parasites, and therefore that $N^*_{\mathrm{VB}}(\phi)$ is defined only for $\phi = \phi_0$. Nevertheless, I include it in all graphs for comparison with that of DT parasites. Also, more generally if we allowed τ to change with transmission mode, and if ϕ changed during an infection for both transmission modes (but did so in different ways for DT and VB parasites), then the $N^*_{\mathrm{DT}}(\phi)$ and $N^*_{\mathrm{VB}}(\phi)$ curves might look different, and they then need not intersect at $\phi = k$. The framework below easily allows for such generality; however, in the absence of clear empirical information about how transmission mode affects these different processes, I employ the simple assumptions outlined above. Moreover, as will be seen below, one of my main findings is that a relatively broad range of evolutionary outcomes is possible (including instances in which Ewald's hypothesis is not valid) even under these simple assumptions. Therefore, allowing for a greater flexibility in the model's assumptions will likely only further this finding.

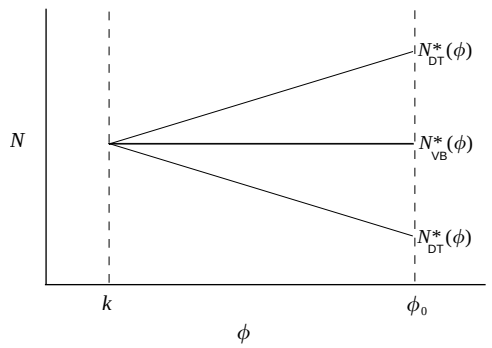

FIG. 3. Plots of the function $N^*(\phi)$ in the $N-\phi$ plane for VB parasites and the two qualitative possibilities for this function for DT parasites. Note that, under the assumptions of Ewald's hypothesis used in the text, $N^*(\phi)$ for VB parasites is independent of ϕ. Indeed, drawing this line in the $N-\phi$ phase space is slightly misleading since it is assumed that ϕ never changes during an infection for VB parasites, and therefore that $N^*_{\mathrm{VB}}(\phi)$ is defined only for $\phi = \phi_0$. Nevertheless, I include it in the graph for comparison with that of DT parasites. More generally, if we allowed τ to change with transmission mode, and if ϕ changed during an infection for both transmission modes (but did so in different ways for DT and VB parasites), then these $N^*(\phi)$ curves might look different, and they then need not intersect at $\phi = k$.

First, suppose $N^*_{\mathrm{DT}}(\phi)$ is increasing with ϕ (i.e., $\mu(N)/\alpha(N)$ is increasing with N) (Fig. 4). In this case we have the $N^*_{\mathrm{DT}}(\phi) > N^*_{\mathrm{DT}}(k) = N^*_{\mathrm{VB}}(\phi)$ for all $\phi \neq k$. Therefore, supposing that both DT and VB parasites begin with the same initial density and contact rate, DT parasites will replicate longer before stopping, and thereby reach higher densities within the host (Fig. 4). This implies

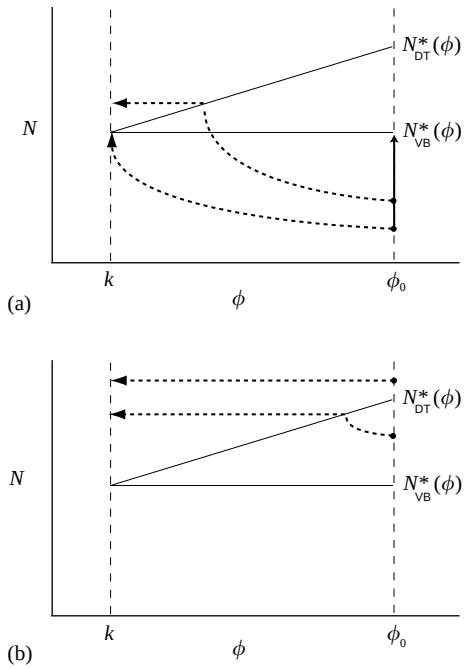

FIG. 4. A comparison of the ES replication strategy of DT versus VB parasites for the case where the function $N^*_{\mathrm{DT}}(\phi)$ increases with ϕ. The thin solid lines represent the functions $N^*_{\mathrm{DT}}(\phi)$ and $N^*_{\mathrm{VB}}(\phi)$ ($N^*_{\mathrm{DT}}(\phi)$ is shown as a straight line for simplicity, but it general it will be some increasing curve). Thick solid arrows represent trajectories in the $N-\phi$ plane under the optimal replication strategy for VB parasites. Thick dotted arrows represent different trajectories in the $N-\phi$ plane under the optimal replication strategy for DT parasites. Trajectories for each parasite type can be deduced using the general qualitative results presented in Figs. 1 and 2. (a) Two examples of initial points $\{N,\phi\}$ that lie below $N^*_{\mathrm{VB}}(\phi)$. (b) Two examples of initial points $\{N,\phi\}$ that lie above $N^*_{\mathrm{VB}}(\phi)$. Note that in (b) the VB parasite never replicates, and thus remains at its starting density.

that, all else equal, DT parasites will result in a larger parasite-induced mortality rate than VB parasites. Notably, this is the opposite of Ewald's prediction.

Now suppose $N^*_{DT}(\phi)$ is decreasing with ϕ (i.e., $\mu(N)/\alpha(N)$ is decreasing with N) (Fig. 5). In this case we have the $N^*_{DT}(\phi) < N^*_{DT}(k) = N^*_{VB}(\phi)$ for all $\phi \neq k$. Therefore, supposing that both DT and VB parasites begin with the same initial density and contact rate, VB parasites will replicate longer before intersecting the $N^*_{VB}(\phi)$ curve than will DT parasites before they intersect the $N^*_{DT}(\phi)$ curve (Fig. 5). Of course, a singular control then occurs for DT parasites, such that $r^* \neq 0$, and the DT parasite continues replicating at a slower (and decreasing) rate until $\phi \to k$, at which point the density approaches $N^*_{DT}(k) = N^*_{VB}(\phi)$. In other words, both types of parasite begin replicating at maximal speed, and both types approach a density within the host of $N^*_{DT}(k)$ (or equivalently $N^*_{VB}(\phi)$) given the infection lasts long enough, but DT parasites switch from a period of maximal growth rate to some intermediate growth rate before reaching this density asymptotically, whereas VB parasites replicate at maximal speed all the way up until this density is reached (in finite time). Therefore, we expect VB parasites to induce a larger total amount of mortality on the host in this case. This is in accord with Ewald's hypothesis, however, it is not clear how substantial this difference will be. The region of time in the infection during which DT parasites have a sub-maximal replication rate can be small, and additionally, it occurs after the infection has been around for a while. If the point at which this switch in replication rate occurs is quite late in an infection, then the infection might well end due to host mortality (or parasite clearance through some form of host defense) prior to this, making this difference in parasite-induced mortality insignificant on average.

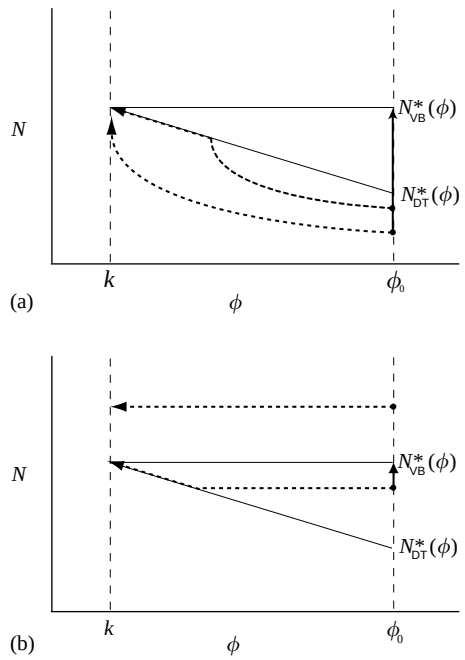

FIG. 5. A comparison of the ES replication strategy of DT versus VB parasites for the case where the function $N^*_{DT}(\phi)$ decreases with ϕ. The thin solid lines represent the functions $N^*_{DT}(\phi)$ and $N^*_{VB}(\phi)$ ($N^*_{DT}(\phi)$ is shown as a straight line for simplicity, but in general it will be some increasing curve). Thick solid arrows represent trajectories in the $N - \phi$ plane under the optimal replication strategy for VB parasites. Thick dotted arrows represent different trajectories in the $N - \phi$ plane under the optimal replication strategy for DT parasites. Trajectories for each parasite type can be deduced using the general qualitative results presented in Figs. 1 and 2. (a) Two examples of initial points $\{N, \phi\}$ that lie below $N^*_{DT}(\phi)$. (b) Two examples of initial points $\{N, \phi\}$ that lie above $N^*_{DT}(\phi)$.

DISCUSSION

The results presented here provide only qualified support for Ewald's hypothesis as a general evolutionary explanation for the higher virulence observed in vector-borne versus directly transmitted parasites. In fact, the first model presented, in which the host contact rate is instantaneously tied to parasite density within host, is the only model examined that supports this hypothesis unequivocally. The second model, which is perhaps more realistic, supports this hypothesis only under certain assumptions and parameter values.

To better understand these conclusions, it is helpful to consider two main selective factors that affect the evolution of virulence under different transmission modes. The first of these is simply the additional cost of virulence to directly transmitted parasites that Ewald identified in his verbal models. Because higher parasite virulence is tied to a reduced host activity level, and because this translates into a reduced transmission potential for DT parasites, this generates a selective advantage for lower virulence in DT parasites (but not VB parasites).

The second factor is related to the timing of parasite transmission. As already discussed, the host activity level will drop during an infection and this translates into a reduced transmission rate between hosts for DT parasites (but not for VB parasites). In addition to imposing selection for reduced virulence for the reasons

just described, this also curtails the effective lifespan of an infection in much the same way that an increase in disease-independent host mortality does (Anderson and May, 1982; Sasaki and Iwasa, 1991; Kakehashi and Yoshinaga, 1992; Lenski and May, 1994; Ebert and Weisser, 1997; Williams and Day, 2001). This can generate a selective advantage for higher replication and thereby higher virulence for DT parasites because it devalues future reproductive output. In other words, DT parasites are expected to have reduced transmission potential relative to VB parasites regardless of the level of virulence, and this favors higher levels of replication.

Which of the two selective factors is the strongest depends upon the assumptions of the model. It was shown that a prediction exactly opposite to Ewald's hypothesis holds (i.e., DT parasites should be more virulent than VB parasites) whenever $\mu(N)/\alpha(N)$ increases with N (where μ is the mortality rate and α is the rate of decrease in contact rate). Since both μ and α are assumed to increase with parasite density, N, this occurs whenever the effect of parasite density on host mortality is stronger than on the rate of decrease in host contact rate. On the other hand, if the reverse holds (i.e., $\mu(N)/\alpha(N)$ decreases with N), so that the effect of parasite density on host mortality is weaker than on the rate of decrease in host contact rate, Ewald's hypothesis is supported. Which of these situation is most reasonable is an empirical issue, but it is questionable whether such specific assumptions are likely to provide a general explanation for observed patterns on virulence. Additionally, the analysis of this case suggests that the differences in virulence between DT and VB parasites might often be quite small. Both DT and VB parasites are predicted to eventually reach the same density within the host, and they differ only in that DT parasites are expected to slow their replication rate late in an infection and reach this density asymptotically, whereas VB parasites reach this density as fast as possible (Figs. 4 and 5).

Although these results provide only qualified support for Ewald's hypothesis, it is conceivable that alternative model formulations might provide more support. Therefore, it is worthwhile considering the limitations of the various simplifications employed here. Perhaps one of the most glaring omissions is the lack of an explicit accounting of the dynamics of an immune response. Such effects can have a large qualitative impact on predictions about virulence evolution (Day, in press). Nevertheless, there are two main justifications for this simplification. First, Ewald's verbal explanation does not explicitly involve any important effects of an immune response, and therefore it is important to neglect this component of the host–parasite interaction as a first attempt in exploring the validity of his prediction. Second, previous theoretical results have been published that employ this same assumption, and therefore using it here allows a more straightforward comparison with previous work. Moreover, given this is the first detailed examination of Ewald's hypothesis, it seems reasonable to err on the side of simplicity for the sake of clarity. Nevertheless, including the dynamics of host defense mechanisms is an important subject for future research.

Another important assumption is that the contact rate between hosts for VB parasites remains constant during an infection. This is unlikely to be strictly true, because changes in host activity level will likely impose some effect on the ability of vectors to transmit the parasite. Indeed, Ewald (1994) has suggested that host immobility as a result of high parasite replication might even increase the contact rate for mosquito-borne parasites. In either case, it is easy to see that relaxing this assumption does not alter the qualitative conclusions for the model where ϕ depends directly on N, provided that $\phi(N)$ decreases more quickly for DT than for VB parasites (see inequality (20)). In fact, to some extent, the same is true for the model in which ϕ depends upon N indirectly. Of course, in this latter case, Ewald's hypothesis has only limited support anyhow, and allowing greater flexibility in the specification of ϕ does not alter this finding.

It is interesting to compare the conclusions derived here with that part of Day (2001) that explores the validity of Ewald's hypothesis. Day (2001) considered the two forms of dependence of contact rate on parasite density that are considered here (i.e., direct and indirect dependence) but made the simplifying assumption that this density must be "chosen" at the beginning of an infection and that it never changes throughout the infection. The present results illustrate that this restrictive assumption does not alter the conclusion that Ewald's hypothesis is not, in general, valid. In fact, the present results reveal that relaxing this assumption and allowing the parasite density within the host to change over an infection makes the validity of Ewald's hypothesis even more restrictive. In both models, Ewald's hypothesis can be valid only if $\mu(N)/\alpha(N)$ decreases with N. But while this is sufficient for there to be a difference in virulence between DT and VB parasites if N is forced to remain constant (Day, 2001), the present results reveal that further assumptions are required if this is to result in substantial virulence differences. The reason is that, because the replication

schedule is evolutionarily flexible here, both VB and DT parasites are predicted to eventually attain the same density within a host. The difference between the two lies solely in the speed with which each type attains this density (Fig. 5). Because the predicted differences in replication rate of DT versus VB parasites often occur late during an infection, this can translate into very little difference is total parasite-induced mortality between the two.

Finally, given that the present results cast some doubt on Ewald's hypothesis as a general explanation for observed patterns of virulence, it is worth considering possible alternative explanations. Boots and Sasaki (1999) offer one possibility based on differences in the ability of VB and DT parasites to exploit hosts in a spatially distributed population. Another alternative, suggested by the model presented here, arises from differences in inoculum size. If VB parasites typically have a larger inoculum size than DT parasites, then all else equal, we might expect them to induce a greater degree of host mortality (Fig. 6). In this case, the VB trajectory starts at a higher density than the DT trajectory. As a result, even if DT parasites are predicted to reach a higher final density than VB parasites (e.g., Fig. 6a), their density will still be lower for some initial period of time during the infection. This can translate into an equal or greater total mortality induced by VB parasites over the course of the infection for a wide variety of functional forms for $N^*_{DT}(\phi)$ and $N^*_{VB}(\phi)$. Additionally, the high density of VB parasites that occurs early in an infection is likely to induce greater total host mortality than the high density of DT parasites that occurs late in an infection because this high density occurring late will often never be realized. In particular, if the host dies from causes unrelated to infection, or if it clears the infection through a defense mechanism, then DT parasites might rarely have the time required to "catch up" to VB parasites in density. Exploring the details of this sort of interaction between host defenses and inoculum size would require a more sophisticated model than that presented here, however, and is an interesting area for future research.

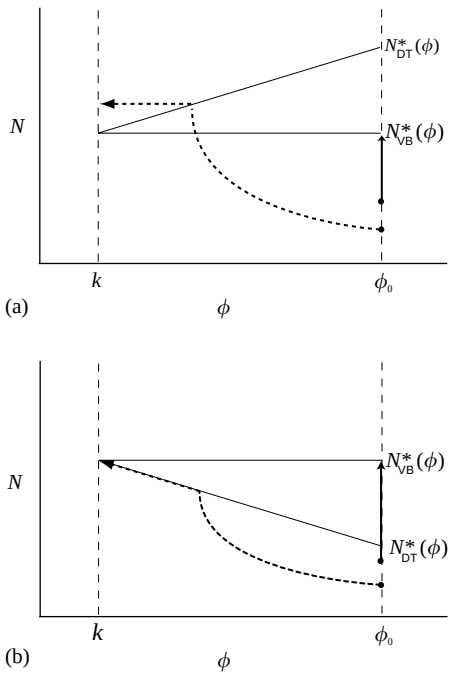

FIG. 6. A depiction of how inoculum size can affect the optimal replication strategy, and thereby the level of virulence. Thin solid lines represent $N^*_{DT}(\phi)$ and $N^*_{VB}(\phi)$. Thick solid arrows represent trajectories under the optimal replication strategy for VB parasites. Thick dotted arrows represent trajectories under the optimal replication strategy for DT parasites. (a) An example where $N^*_{DT}(\phi)$ increases with ϕ—in this case the within host density (and therefore the virulence) of VB parasites is greater than that for DT parasites for some period of time at the beginning of the infection since it starts at a higher initial density. (b) An example where $N^*_{DT}(\phi)$ decreases with ϕ. In this case the within host density (and therefore the virulence) of VB parasites is always greater than that for DT parasites.

APPENDIX A

Here I derive the conditions that the ES replication strategy must satisfy. A related derivation can be found in Sasaki and Iwasa (1991). I characterize the optimal replication strategy by deriving conditions that it must satisfy as a function of the parasite density within the host.

From Eq. (10) we can see that, because r enters this equation linearly, r^* must take either of the extreme values of $r^* = 0$ or $r^* = r_{max}$ depending upon the sign of dV^*/dN. Technically, one must also consider the possibility that r^* takes on intermediate values if $dV^*/dN = 0$ over some interval of time (i.e., if there is a "singular control"; e.g., see Appendix C), but this is not possible in the current model. In particular, it can be shown that if $dV^*/dN = 0$ over some interval of time, then $r^* = 0$ over this interval. To prove this, suppose that $dV^*/dN = 0$ over some interval of time. In this case, Eq. (10) reveals that

$$V^* = \frac{\beta(N)}{\delta + bN}; \tag{A1}$$

(note that this satisfies the boundary conditions $V^*(0) = 0$ and $\lim_{N\to\infty} V^*(N) = 0$ since $\beta(0) = 0$ and β never exceeds some maximum value for all N). But this expression for V^* is exactly the definition of $W(N)$ given in Eq. (11), and therefore having $dV^*/dN = 0$ requires having $dW/dN = 0$. Since there is assumed to be a single value of N for which $dW/dN = 0$ (i.e., $N = N^*$), we require that $N = N^*$ over this interval of time, which thereby requires that N be unchanging; i.e., $r^* = 0$. Therefore, r^* must always be equal to 0 or $r_{\max}$ at all times.

Now suppose that $N < N^*$. It can be proven, by contradiction, that r^* must equal $r_{\max}$ in this case. In particular, suppose that $r^* = 0$ when $N < N^*$. From Eq. (10) we can see that Eq. (A1) must then hold for V^*, and we also require that $dV^*/dN \leqslant 0$. But this implies that $dW/dN \leqslant 0$ which, itself, implies that $N \geqslant N^*$ resulting in a contradiction. Therefore, since r^* cannot be zero when $N < N^*$, it must be equal to $r_{\max}$.

Now I will prove that $r^* = 0$ if $N \geqslant N^*$, by showing that all other replication schedules result in lower fitness. First notice that, because β reaches its maximal value for some finite value of N, eventually we must have $r^* = 0$ from some point in the infection onward, since at very least, once N is greater than the value at which $\beta = \beta_{\max}$, increased replication results in additional mortality costs without any transmission benefit (note that this argument relies on the assumption that, if r^* were not eventually zero from some time onward, N would continue to increase without bound. Technically this need not be the case. For example, N might remain bounded if r^* is made up of an infinite sequence of alternating intervals with $r^* = 0$ and $r_{\max}$, where the duration of each interval decreased to zero. It is possible, however, to show that this is not an optimal strategy; unpubl. results).

Now suppose that we have an optimal replication schedule, $r_1(t)$, that is not identically equal to zero for all $N > N^*$, and let $\hat{t}$ be the time at which $r_1(t) = 0$ for all $t \in (\hat{t}, \infty)$; i.e., it is the time at which this replication schedule does finally switch to non-growth (we know that any schedule that is optimal must have this property). The reproductive value a small interval of time before this stage is then given by

$$V(N(\hat{t}-\Delta t); r_1) = \frac{1}{l(\hat{t}-\Delta t)} \int_{\hat{t}-\Delta t}^{\infty} \beta(N) l(\hat{t}-\Delta t) \times \exp\left\{ -\int_{\hat{t}-\Delta t}^{t} (\delta + bN(s))\, \mathrm{d}s \right\} \mathrm{d}t \tag{A2}$$

$$\begin{aligned}
&= \frac{1}{l(\hat{t}-\Delta t)} \left\{ \begin{array}{l} \int_{\hat{t}-\Delta t}^{\hat{t}} \beta(N) l(\hat{t}-\Delta t) \\ \quad \times \exp\{ -\int_{\hat{t}-\Delta t}^{t} (\delta + bN(s))\, \mathrm{d}s \}\, \mathrm{d}t \\ \quad + \int_{\hat{t}}^{\infty} \beta(N) l(\hat{t}-\Delta t) \\ \quad \times \exp\{ -\int_{\hat{t}-\Delta t}^{t} (\delta + bN(s))\, \mathrm{d}s \}\, \mathrm{d}t \end{array} \right\} \\
&\approx \beta(N(\hat{t}-\Delta t))\Delta t \\
&\quad + \frac{l(\hat{t})}{l(\hat{t}-\Delta t)} \frac{1}{l(\hat{t})} \\
&\quad \times \int_{\hat{t}}^{\infty} \beta(N) \exp\left\{ -\int_{0}^{t} (\delta + bN(s))\, \mathrm{d}s \right\} \mathrm{d}t \\
&\approx \beta(N(\hat{t}-\Delta t))\Delta t \\
&\quad + [1 - (\delta + bN(\hat{t}-\Delta t))\Delta t] V(N(\hat{t}); r_1) \\
&= \beta(N(\hat{t}-\Delta t))\Delta t \\
&\quad + [1 - (\delta + bN(\hat{t}-\Delta t))\Delta t] W(N(\hat{t}))
\end{aligned} \tag{A3}$$

for small Δt, where the last equation follows from the fact that N is constant from time $\hat{t}$ onwards, and therefore V as a function of N is identical to $W(N)$ (and is given by Eq. (A1)). On the other hand, consider a replication schedule, r_2, such that $r_2(t) = 0$ for all $t \in (\hat{t}-\Delta t, \infty)$, but is otherwise identical to r_1; i.e., it stops replication Δt units of time earlier. By a similar derivation, its reproductive value at $\hat{t}-\Delta t$ is

$$\begin{aligned}
&V(N(\hat{t}-\Delta t); r_2) \\
&\quad \approx \beta(N(\hat{t}-\Delta t))\Delta t \\
&\quad + [1 - (\delta + bN(\hat{t}-\Delta t))\Delta t] W(N(\hat{t}-\Delta t))
\end{aligned} \tag{A4}$$

for small Δt. The only difference between Eqs. (A3) and (A4) is that the W in Eq. (A4) is evaluated at $N(\hat{t}-\Delta t)$ rather than at $N(\hat{t})$ because that strain stopped replicating at time $\hat{t}-\Delta t$. Therefore, since

$$\begin{aligned}
&V(N(\hat{t}-\Delta t); r_1) - V(N(\hat{t}-\Delta t); r_2) \\
&\quad = [1 - (\delta + bN(\hat{t}-\Delta t))\Delta t] \\
&\quad \times \{ W(N(\hat{t})) - W(N(\hat{t}-\Delta t)) \},
\end{aligned} \tag{A5}$$

which is negative provided both $N(\hat{t})$ and $N(\hat{t}-\Delta t)$ are larger than N^*, we can see that the fitness of r_2 is greater than that of r_1; i.e., it would have been better for the parasite to stop replicating a small amount of time earlier. This argument can be carried out iteratively, backwards in time to demonstrate that the optimal replication strategy must have $r^* = 0$ as long as $N \geqslant N^*$.

APPENDIX B

To determine the sign of $dN^*/d\phi$, I implicitly differentiate Eq. (29) with respect to ϕ to obtain

$$\frac{\partial^2 W}{\partial N^2}\frac{dN^*}{d\phi} + \frac{\partial^2 W}{\partial N \partial \phi} = 0. \quad \text{(B1)}$$

Because I am assuming that values of N that satisfy Eq. (29) are maxima (with respect to N), we have that $\partial^2 W/\partial N^2 < 0$, and therefore Eq. (B1) reveals that

$$\frac{dN^*}{d\phi} \propto \frac{\partial^2 W}{\partial N \partial \phi}. \quad \text{(B2)}$$

Using the fact that we also have $\partial W/\partial N = 0$ when Eq. (B1) holds, a bit of manipulation then demonstrates that

$$\frac{dN^*}{d\phi} \propto \frac{d}{dN}\frac{\mu(N)}{\alpha(N)}. \quad \text{(B3)}$$

Thus, the curve $N^*(\phi)$ increases (decreases) with ϕ if μ/α increases (decreases) with N.

APPENDIX C

Here I derive the conditions that the ES replication strategy must satisfy for the case where ϕ depends indirectly on N. I will treat the case where $N^*(\phi)$ is an increasing function of ϕ separately from the case where it is a decreasing function of ϕ. I characterize the optimal replication strategy by deriving conditions that it must satisfy as a function of the parasite density within the host and the contact rate; i.e., $\{N, \phi\}$. Notice that N can take on any non-negative value whereas we are interested only in values of ϕ that satisfy $k \leqslant \phi \leqslant \phi_0$.

$N^*(\phi)$ *Increases with* ϕ

From Eq. (27) we can see that r enters linearly, and therefore, an intermediate value of r^* does not occur (except possibly during a period of "singular control"). In other words, r^* must take either of the extreme values of $r^* = 0$ or $r^* = r_{\max}$ depending upon the sign of $\partial V^*/\partial N$. In fact, in this case it can be proven that it is never possible for $\partial V^*/\partial N = 0$ over some interval of time, and therefore it is never possible for there to be a singular control. In particular, if $\partial V^*/\partial N = 0$, then Eq. (27) becomes

$$0 = \tau(N)\phi - (\delta + bN)V^* - \frac{\partial V^*}{\partial \phi}\alpha(N)(\phi - k), \quad \text{(C1)}$$

with the additional boundary condition $V^*(N, k) = \tau k/\mu$. This can be solve for V^* to give

$$V^*(N, \phi) = \frac{\tau(N)\phi}{\mu(N)}\frac{(k/\phi)\alpha(N) + \mu(N)}{\alpha(N) + \mu(N)}. \quad \text{(C2)}$$

Notice that this expression for V^* is exactly the definition of $W(N, \phi)$ given in Eq. (28). Therefore, for a singular control we must have $\partial W/\partial N = 0$ over some interval of time. However, this quantity changes through time according to

$$\frac{d}{dt}\frac{\partial W}{\partial N} = \frac{\partial^2 W}{\partial N^2}\frac{dN}{dt} + \frac{\partial^2 W}{\partial \phi \partial N}\frac{d\phi}{dt} \quad \text{(C3)}$$

$$= \frac{\partial^2 W}{\partial N^2} r^* N - \frac{\partial^2 W}{\partial \phi \partial N}\alpha(N)(\phi - k). \quad \text{(C4)}$$

The first term of Eq. (C4) is non-positive, and the fact that $N^*(\phi)$ is an increasing function of ϕ means that $\partial^2 W/\partial \phi \partial N > 0$. Therefore, unless $\phi = k$ and/or $\alpha \equiv 0$, expression (C4) will be negative, indicating that $\partial W/\partial N$ must change though time and therefore cannot equal 0 except for an isolated point during the infection. Therefore, either $r^* = 0$ and $r_{\max}$ at all times during the infection.

Now let's begin by supposing that the point $\{N, \phi\}$ lies below that function $N^*(\phi)$ in the $N - \phi$ plane (Fig. 4). In other words, the parasite density is less than $N^*(\phi)$ for its corresponding value of ϕ. In this case it can be proven, by contradiction, that we must have $r^* = r_{\max}$. To do so, suppose that $r^* = 0$. In this case Eq. (27) again becomes Eq. (C1), yielding Eq. (C2) as the definition of V^*. Now if $r^* = 0$, then we must also have that $\partial V^*/\partial N < 0$ in Eq. (27). But this is true only for points lying above the function $N^*(\phi)$ in the $N - \phi$ plane. Therefore, we cannot have $r^* = 0$; instead we must have $r^* = r_{\max}$.

There is no need to consider points in the $N - \phi$ plane that lie on the function $N^*(\phi)$ because they can only do so only for an isolated point in time. Now focus attention on that situation in which the point $\{N, \phi\}$ lies above the function $N^*(\phi)$. Clearly, once the infection is such that the parasite density and contact rate lie above the function $N^*(\phi)$, it will do so for all future time since N is non-decreasing and ϕ is non-increasing. In this case is can be proven that $r^* = 0$ following an approach based on that used in Appendix A.

Because $\tau(N)$ reaches its maximal value for some finite value of N, eventually we must have $r^* = 0$ because once N is greater than the value at which $\tau = \tau_{\max}$, increased replication results in additional mortality costs and a faster decay in the contact rate (which also lowers fitness) without producing any transmission benefit. Now suppose that we have an optimal replication schedule, $r_1(t)$, that is not identically equal to zero, and let $\hat{t}$ be the time at which $r_1(t) = 0$ for all $t \in (\hat{t}, \infty)$. The reproductive value a small interval of time before

this stage can be derived as in Appendix A to give

$$\begin{aligned} & V(N(\hat{t}-\Delta t),\phi(\hat{t}-\Delta t);r_1) \\ & \approx \tau(N(\hat{t}-\Delta t))\phi(\hat{t}-\Delta t)\Delta t \\ & + [1-(\delta+bN(\hat{t}-\Delta t))\Delta t]W(N(\hat{t}),\phi(\hat{t})) \end{aligned} \quad \text{(C5)}$$

for small Δt, where $W(N,\phi)$ is given by Eq. (C2). On the other hand, consider a replication schedule, r_2, such that $r_2(t)=0$ for all $t \in (\hat{t}-\Delta t,\infty)$, but is otherwise identical to r_1; i.e., it stops replication Δt units of time earlier. By a similar derivation, its reproductive value at $\hat{t}-\Delta t$ is

$$\begin{aligned} & V(N(\hat{t}-\Delta t),\phi(\hat{t}-\Delta t);r_2) \\ & \approx \tau(N(\hat{t}-\Delta t))\phi(\hat{t}-\Delta t)\Delta t \\ & + [1-(\delta+bN(\hat{t}-\Delta t))\Delta t] \\ & W(N(\hat{t}-\Delta t),\phi(\hat{t}-\Delta t)) \end{aligned} \quad \text{(C6)}$$

for small Δt. The only difference between Eqs. (C5) and (C6) is that the W in Eq. (C6) is evaluated at $\{N\times(\hat{t}-\Delta t),\phi(\hat{t}-\Delta t)\}$ rather than at $\{N(\hat{t}),\phi(\hat{t})\}$ because that strain stopped replicating at time $\hat{t}-\Delta t$. Therefore, since

$$\begin{aligned} & V(N(\hat{t}-\Delta t),\phi(\hat{t}-\Delta t);r_1) \\ & - V(N(\hat{t}-\Delta t),\phi(\hat{t}-\Delta t);r_2) \\ & = [1-(\delta+bN(\hat{t}-\Delta t))\delta t] \\ & \quad \{W(N(\hat{t}),\phi(\hat{t}))-W(N(\hat{t}-\Delta t),\phi(\hat{t}-\Delta t))\}, \end{aligned} \quad \text{(C7)}$$

which is negative provided both $\{N(\hat{t}),\phi(\hat{t})\}$ and $\{N\times(\hat{t}-\Delta t),\phi(\hat{t}-\Delta t)\}$ lie above $N^*(\phi)$, we can see that the fitness of r_2 is greater than that of r_1; i.e., it would have been better for the parasite to stop replicating a small amount of time earlier. Again, this argument can be repeated backwards in time to demonstrate that the optimal replication strategy must have $r^*=0$ as long as $\{N,\phi\}$ lies above $N^*(\phi)$.

$N^*(\phi)$ *Decreases with* ϕ

The main qualitative difference between this case and that examined previously is the possibility of a singular control during which we have $0 \leqslant r^* \leqslant r_{\max}$. As derived previously, for this to occur requires that Eq. (C4) equal zero for some interval of time. In such case, r^* must satisfy

$$r^* = \frac{\partial^2 W}{\partial\phi\partial N}\alpha(N)(\phi-k)\Big/\frac{\partial^2 W}{\partial N^2}N. \quad \text{(C8)}$$

This is clearly feasible since the numerator of (C8) is non-positive and the denominator is negative (Appendix B). Importantly, however, this singular control can occur, only along the $N^*(\phi)$ curve. The reason is that, at a singular control we have $\partial V^*/\partial N = 0$, which implies that V^* is given by Eq. (C2). Since we must also have $\partial V^*/\partial N = 0$ in this case, only points along the curve defined by $N^*(\phi)$ will satisfy both conditions. Notice, however, that if $r_{\max}$ is very small, then even if $r^*=r_{\max}$ it might not be possible for the trajectory in the $N-\phi$ plane to follow the curve $N^*(\phi)$ because N might not increase quickly enough. Throughout I will assume that $r_{\max}$ is large enough so that this is not the case.

Now by an argument similar to that above, it can be proven that r^* cannot equal zero for points $\{N,\phi\}$ below the $N^*(\phi)$ curve. Because a singular control can occur only *on* the $N^*(\phi)$ curve, we must therefore have $r^*=r_{\max}$ in this region.

To analyze the optimal replication strategy when $\{N,\phi\}$ lies above $N^*(\phi)$, I further subdivide this region into two sets: $A \equiv \{\{N,\phi\}: N \geqslant N^*(k)\}$ and $B \equiv \{\{N,\phi\}: N < N^*(k)$ and $\{N,\phi\}$ lies above $N^*(\phi)\}$ (Fig. 5). If $\{N,\phi\} \in A$, then, because we must eventually have $r^*=0$, a chain of reasoning identical to that used above demonstrates that a replication schedule having $r^*=0$ for all time results in the largest fitness.

Now suppose $\{N,\phi\} \in B$. Again a chain of reasoning similar to that used for region A demonstrates that the optimal replication schedule is one in which $r^*=0$ until the trajectory intersects with $N^*(\phi)$. At this point, a singular control occurs. The optimal strategy then results in the trajectory remaining on this curve forever, and as $\phi \to k$, we have $r^* \to 0$.

ACKNOWLEDGMENTS

I thank two anonymous reviewers for several helpful suggestions. This research was supported by a grant from the Natural Science and Engineering Research Council of Canada and a Premier's Research Excellence Award to T.D.

REFERENCES

Anderson, R. M., and May, R. M. 1982. Coevolution of hosts and parasites, *Parasitology* **85**, 411–426.

Anderson, R. M., and May, R. M. 1991. "Infectious Diseases of Humans: Dynamics and Control," Oxford Univ. Press, Oxford, UK.

Antia, R, Levin, B. R., and May, R. M. 1994. Within-host population dynamics and the evolution and maintenance of microparasite virulence, *Am. Nat.* **144**, 457–472.

Boots, M., and Sasaki, A., 1999. Small worlds and the evolution of virulence: Infection occurs locally and at a distance, *Proc. R. Soc. Lond. B* **266**, 1933–1938.

Bremermann, H. J., and Pickering, J., 1983. A game-theoretical model of parasite virulence, *J. Theor. Biol.* **100**, 411–426.

Bull, J. J. 1994. Virulence, *Evolution* **48**, 1423–1437.

Day, T. 2001. Parasite transmission modes and the evolution of virulence, *Evolution* **55**, 2389–2400.

Day, T. On the evolution of virulence and the relationship between various measures of mortality. *Proc. R. Soc. Lond. B*, in press.

Diekmann, O., and Heesterbeek, J. A. P. 2000. "Mathematical Epidemiology of Infectious Disease," Wiley, New York.

Diekmann, O., Heesterbeek, J. A. P., and Metz, J. A. J. 1990. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations, *J. Math. Biol.* **28**, 365–382.

Ebert, D. 1999. The evolution and expression of parasite virulence, *in* "Evolution in Health and Disease" (S. C. Stearns, Ed.), pp. 161–172, Oxford Univ. Press, Oxford, UK.

Ebert, D., and Herre, E. A., 1996. The evolution of parasitic diseases, *Parasitol. Today* **12**, 98–101.

Ebert, D., and Weisser, W. W. 1997. Optimal killing for obligate killers: The evolution of life histories and virulence in semelparous parasites. *Proc. R. Soc. Lond B*, **264**, 985–991.

Ewald, P. W. 1983. Host–parasite relations, vectors, and the evolution of disease severity, *Ann. Rev. Ecol. Syst.* **14**, 465–485.

Ewald, P. W. 1991. Waterborne transmission and the evolution of virulence among gastrointestinal bacteria, *Epidemiol. Infect.* **106**, 83–119.

Ewald, P. W. 1994. "Evolution of Infectious Diseases," Oxford Univ. Press, Oxford, UK.

Ewald, P. W. 1995. Response (to van Baalen and Sabelis), *Trends Microbiol.* **3**, 416–417.

Frank, S. A. 1996. Models of parasite virulence, *Quart. Rev. Biol.* **71**, 37–78.

Kakehashi, M., and Yoshinaga, F. 1992. Evolution of airborne infectious diseases according to changes in characteristics of the host population, *Ecol. Res.* **7**, 235–243.

Keeling, M. J., and Grenfell, B. T. 2000. Individual-based perspectives on R_0, *J. Theor. Biol.* **203**, 51–61.

Lenski, R. E., and May, R. M. 1994. The evolution of virulence in parasites and pathogens: reconciliation between two competing hypotheses. *J. Theor. Biol.* **169**, 253–266.

Levin, B. R. 1996. The evolution and maintenance of virulence in microparasites. *Emerging Infectious Dis.* **2**, 93–102.

Levin, B. R., Bull, J. J., and Stewart, F. M. 1996. The intrinsic rate of increase of HIV/AIDS: Epidemiological and evolutionary implications, *Math. Biosci.* **132**, 69–96.

May, R. M., and Nowak, M. 1995. Coinfection and the evolution of parasite virulence, *Proc. R. Soc. Lond. B* **261**, 209–215.

Nowak, M., and May, R. M. 1994. Superinfection and the evolution of parasite virulence, *Proc. R. Soc. Lond. B* **255**, 81–89.

Read, A. F. et al. 1999. What can evolutionary biology contribute to understanding virulence? *in* "Evolution in Health and Disease" (S. C. Stearns, Ed.), pp. 205–215, Oxford Univ. Press, Oxford, UK.

Sasaki, A., and Iwasa, Y. 1991. Optimal growth schedule of pathogens within a host: Switching between lytic and latent cycles, *Theor. Popul. Biol.* **39**, 201–239.

Stearns, S. C. (Ed.), 1999. "Evolution in Health and Disease," Oxford Univ. Press, Oxford, UK.

Trevathan, W. R., Smith, E. O., and McKenna, J. J. (Eds.), 1999. "Evolutionary Medicine," Oxford Univ. Press, Oxford, UK.

Van Baalen, M., and Sabelis, M. W. 1995. The scope for virulence management: A comment on Ewald's view on the evolution of virulence, *Trends Microbiol.* **3**, 414–416.

Williams, P. D., and Day, T. 2001. Interactions between sources of mortality and the evolution of parasite virulence, *Proc. R. Soc. London B*, in press.

Williams, G. C., and Nesse, R. M. 1991. The dawn of Darwinian medicine, *Quart. Rev. Biol.* **66**, 1–22.