Highlights

- We investigate the invasion probability of vector-borne diseases with multiple hosts
- We study the role of heterogeneous transmission among hosts using branching processes
- We show that invasion probability via infected vector increases with heterogeneity
- We show that invasion probability via infected host can decrease with heterogeneity
- We find higher risk of outbreaks occurrence when pathogen is introduced by vectors

Title

The role of heterogeneity on the invasion probability of mosquito-borne diseases in multi-host models

Authors

Bolzoni L.[1,2], Pugliese A.[3], Rosà R.[2]

[1] Direzione Sanitaria – Servizio di Analisi del Rischio, Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia Romagna, Parma, Italy.

[2] Department of Biodiversity and Molecular Ecology, Research and Innovation Centre – Fondazione Edmund Mach, San Michele all'Adige (TN), Italy.

[3] Department of Mathematics, University of Trento, Povo (TN), Italy

Corresponding Author

Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia Romagna,
Via dei Mercati 13, 43100 Parma, Italy
e-mail: luca.bolzoni@izsler.it

ABSTRACT

Heterogeneity in transmission and stochastic events can play a significant role in shaping the epidemic dynamics of vector-borne infections, especially in the initial phase of an outbreak. In this work, by using multi-type branching process methodologies, we assess how heterogeneities in transmission among a large number of host groups can affect the invasion probabilities of a mosquito-borne disease.

We show with both analytical and numerical methods that heterogeneities in transmission can shape the invasion probabilities differently from how they affect the basic reproduction number (R_0). In particular, we find that, while R_0 always increases with the heterogeneity, the invasion probability after the introduction of infected hosts can decrease with the increase of transmission heterogeneity, even approaching zero when the number of host groups is very large. In addition, we show that the invasion probability via infected vectors is always larger than via infected hosts when heterogeneous transmission is sufficiently high.

Our findings suggest that, for multi-species infections (e.g. West Nile fever and Rift Valley fever) or for single-species infections with patchy host distribution, the introduction of primary infected vectors may represent a higher risk for major outbreaks occurrence than introductions of infected hosts.

KEYWORDS: Vector-borne infection; Branching process; Multiple host; Heterogeneity; Stochasticity; Multi-group model

1. INTRODUCTION

Vector-borne diseases are infections transmitted by the bite of infected arthropods, such as mosquitoes, ticks and fleas. Among them, mosquito-borne diseases (such as malaria, West Nile virus, chikungunya, dengue fever, Rift Valley fever and yellow fever) represent major threats to human and animal health.

Mathematical models have been widely developed aiming at describing the complexity of host-mosquito-pathogen interactions. The first mathematical description of mosquito-borne infections is due to Sir Ronald Ross, who provided a synthetic theoretical framework for the transmission of human malaria (Ross 1911). His pioneering work, later extended by Macdonald (1952), provided several insight on vector-borne disease control and prevention. The main achievement of the Ross-Macdonald model was the identification of a threshold condition for disease invasion – the basic reproduction number – which still is, to this day, the most important metric in mathematical epidemiology (see Section 2.1 for more details).

In more recent years the basic theory of the Ross-Macdonald model has been expanded to include eco-epidemiological complexities inherent in malaria and other mosquito-borne infections, such as waning immunity (Aron 1988), multiple strain co-circulation (Ferguson et al. 1999), cross-immunity (Adams et al. 2006), seasonal variations (Dietz 1971), and spatial, behavioural or genetic heterogeneity in transmission (Woolhouse et al. 1997). Although this large amount of work significantly improved the understanding of vector-borne diseases epidemiology, two recent review articles on mathematical models of mosquito-borne diseases published by Reiner et al. (2013) and Smith et al. (2014) pointed out that there exist still a small number of studies looking at the consequences of poorly mixed vector-host encounter and spatial heterogeneity on disease dynamics. Several empirical surveys provided evidence for the existence of heterogeneities in the frequency of mosquito bites among humans. Differences in mosquito biting rate have been observed among people with different blood group type (Shirai et al. 2004), carbon dioxide emission – hence size – (Zainulabeuddin & Leal 2007), health status of both vector and host (Ferguson & Read 2004), skin

bacteria composition (Verhulst et al. 2011), etc. Similarly, heterogeneous biting rates have been observed among the avian community of West Nile virus host species both in USA (Kilpatrick et al. 2006a) and Europe (Roiz et al. 2012).

From the epidemiological point of view, seminal theoretical papers by Barbour (1978), Dye & Hasibeder (1986) and Hasibeder & Dye (1988) based on deterministic models showed that heterogeneities in the transmission of vector-borne infections play a major role in defining the threshold condition for disease invasion. In particular, they show that, when vector biting rate differs among host groups, the basic reproduction number is always larger than under the hypothesis of completely homogeneous mixing.

However, deterministic models ignore the contribution of demographic stochasticity, which is especially relevant when the prevalence in hosts and/or vectors is low (for instance at the beginning of an outbreak). Historically, stochastic epidemic models have been developed within the framework of continuous-time Markov processes or branching processes at the population level (Bartlett 1964; Bailey 1975), although more complex models taking into account many details of mosquito-borne infections have been recently proposed (Magori et al. 2009; Perkins et al. 2013). The development of the theory of stochastic epidemic models has allowed, largely on the basis of branching process approximations, for the computation of disease invasion probabilities in several types of models (Andersson & Britton 2000). In the case of vector-borne diseases, branching process approximations have been extensively analysed in the case of one host (Bartlett 1964; Griffiths 1972; Ball 1983; Lloyd et al. 2007) and two host groups (Lloyd et al. 2007). Here, we extend the analysis to the case of *n* different host groups by studying the effect of stochastic processes on the invasion probability of a vector-borne infection under different assumptions of heterogeneous host-vector mixing. In particular, we analyse the role of the number of host groups (Subsection 3.3.1.) and of the heterogeneity in vector biting rates among them (Subsection 3.3.2.) in shaping the invasion probability of infections introduced in a new population by a single vector or host.

2. THE DETERMINISTIC MULTI-GROUP MODEL

The multi-group model (Dye & Hasibeder 1986) is a generalization of the classical Ross-Macdonald host-vector model (Ross 1911; Macdonald 1952), which takes into account different host types. Each host type can indicate either a host species (Kilpatrick et al. 2007) or the patch (or group) to which an individual host belongs in populations characterized by spatial/behavioural heterogeneity (Woolhouse et al. 1997). The model assumes that the vector population (V) and the n host type populations (H_j, with $j = 1,\ldots,n$) are constant in their sizes, and that they can be subdivided at any time t into two compartments with respect to the disease: infectious (I for vectors and Y_j for type-j host) or susceptibles ($V - I$ and $H_j - Y_j$, respectively).

A susceptible vector [host] can acquire infection by biting [being bitten by] an infected host [vector]. The multi-group model assumes that the rate at which vectors bite hosts is a constant, say λ, independent of host density. This implies that host density does not represent a limiting factor for vectors to find a valuable meal. Then, the rate at which susceptible vectors become infected is equal to λ, times the probability that the bite is on an infected host, times the probability that the vector becomes infected (assumed to be a constant value q_V independent of host type).

Letting γ_j represent the proportion of vector bites allocated to type-j host (with $\sum_j \gamma_j = 1$) and δ the mortality rate of vectors, one arrives at the equation

$$\dot{I} = \beta(V - I)\left(\sum_{j=1}^{n} \gamma_j \frac{Y_j}{H_j}\right) - \delta I \qquad (1)$$

where $\beta = \lambda q_V$.

Similarly, a susceptible host of type j gets infected when it is bitten by an infected vector (this occurs at rate $\lambda\gamma_j I[1 - Y_j/H_j]$), times the probability of becoming infected in that case, say q_H. Then, if ζ represents the recovery rate of infected hosts (assumed also to be constant among types), the equations for infected type-j hosts are

$$\dot{Y}_j = \alpha\gamma_j I\left(1-\frac{Y_j}{H_j}\right)-\zeta Y_j, \quad j=1,\ldots,n \tag{2}$$

where $\alpha = \lambda\, q_{H.}$

Equations (1)-(2) constitute the system of differential equations for the multi-group model as in Dye & Hasibeder (1986).

Transmission is considered to be homogeneous if the fraction of bites γ_j allocated to type-*j* hosts is proportional to their relative abundance, i.e. if $\gamma_j = h_j$ (where $h_j = H_j / H$ and $H = \sum_j H_j$ is total host abundance). Correspondingly, the heterogeneity generated by variations among different host types in their exposure to mosquito feeding can be measured by the variance of the γ_j / h_j weighted by the frequency distribution of h_j, namely $\mathrm{var}(\gamma_j / h_j ; h_j) = \sum_j \left[(\gamma_j - h_j)^2 / h_j\right]$.

2.1. The basic reproduction number

The basic reproduction number, R_0, represents the average number of secondary infections that one primary infective individual produces during its infectious period into an entirely susceptible population (Diekmann et al. 1990). In the case of vector-borne diseases the infection cycle lies in a two-step process: from host-to-vector and from vector-to-host. Then, as shown by Dye & Hasibeder (1986), the basic reproduction number for multi-group model (1)-(2) can be broken up in two different terms: $R_0^{H_jV} = \beta\gamma_j V /(\zeta H h_j) = r_{HV}\gamma_j / h_j$, which represents the average number of secondary infections among vectors that arise from a single type-*j* host; and $R_0^{VH_j} = \alpha\gamma_j / \delta = r_{VH}\gamma_j$, which represents the average number of secondary infections among type-*j* hosts that arise from a single vector. Thus, over the entire transmission cycle, one infective host or vector gives rise to an average of

$$R_0 = \sum_{j=1}^{n} R_0^{H_jV} R_0^{VH_j} = \frac{\alpha\beta V}{\zeta\delta H}\sum_{j=1}^{n}\frac{\gamma_j^2}{h_j} = r_{HV} r_{VH}\sum_{j=1}^{n}\frac{\gamma_j^2}{h_j} \tag{3}$$

where $\sum_j \gamma_j^2 / h_j$ is defined as the relative reproduction number, $R_{0,\text{rel}}$, which represents the proportion of R_0 due to the heterogeneous mixing among hosts and vectors (Woolhouse et al. 1997). In particular, Dye & Hasibeder (1986) showed that the relative reproduction number can be rewritten as

$$R_{0,rel} = \sum_{j=1}^{n} \frac{\gamma_j^2}{h_j} = 1 + \text{var}\left(\frac{\gamma_j}{h_j}; h_j \right). \quad (4)$$

Equation (4) shows that heterogeneous vector biting rates and host frequencies in the community, corresponding to $\text{var}(\gamma_j / h_j; h_j) > 0$, lead to a greater R_0 value in (3) than in the case of homogeneous mixing, characterized by $\text{var}(\gamma_j / h_j; h_j) = 0$. In summary, deterministic multi-group model (1)-(2) predicts that both infection invasion and persistence increase as a consequence of non-homogeneous mixing (Dye & Hasibeder 1986).

We stated that the quantity R_0 defined above can be interpreted either as the expected number of infected vectors generated over a cycle, in completely susceptible populations, by one infected vector, or the same quantity in terms of infected hosts. While the former interpretation is straightforward, the latter is subtler and relies on the argument by Diekmann et al. (1990). Indeed, an infected type-*j* host will generate on average $R_0^{H_j V} R_0^{VH_i}$ infectives of type-*i* host for all *i*. Then, we can define a matrix *M*, with $M_{ij} = R_0^{H_j V} R_0^{VH_i}$, describing all average transmissions depending on the type of the initial infected. Following Diekmann et al. (1990), the average number of transmissions from an average infected host, R_0, is defined as the spectral radius of *M*; a simple computation then leads to the expression (3). In this case, it is possible to provide a more intuitive understanding of the quantity R_0 by saying that an initial type-*j* infected host will on average infect M_{ij} type-*i* hosts for a total number of first generation infected (I_1)

$$I_1 = \sum_{i=1}^{n} R_0^{H_j V} R_0^{VH_i} = R_0^{H_j V} \sum_{i=1}^{n} R_0^{VH_i} .$$

Now, each of the type-*i* hosts will infect on average M_{ki} type-*k* hosts for a total number of second generation infected (I_1)

$$I_2 = \sum_{i,k=1}^{n} R_0^{H_j V} R_0^{VH_i} R_0^{H_i V} R_0^{VH_k} = R_0^{H_j V} \sum_{k=1}^{n} R_0^{VH_k} \sum_{i=1}^{n} R_0^{VH_i} R_0^{H_i V} \ .$$

Then, we obtain

$$I_2 / I_1 = \sum_{i=1}^{n} R_0^{H_i V} R_0^{VH_i} = R_0 ,$$

showing that each first generation infected host infects on average R_0 hosts. This argument also shows that we can expect more variability in the outcome when an infected host is introduced than after the introduction of an infected vector. Indeed, the type of initial host influences the expected number of first generation infected.

3. THE INDIVIDUAL-BASED MULTI-GROUP MODEL

Deterministic models are useful tools for understanding pathogen dynamics, however the treatment of host and vector populations as continuously varying quantities can produce unrealistic results. For instance, in deterministic models pathogens never wholly die out and can regenerate from arbitrarily small amounts of residual infection (Mollison 1991). By including demographic stochasticity, we explicitly describe the movements of individuals between classes, instead of considering the average rates at which individuals move (Bartlett 1960). A stochastic description of the infection process allows for the last infected individual to recover or die before the infection is transmitted even in the case R_0 is larger than 1. Thus, differently from the deterministic model, infection can go extinct as a consequence of stochastic events and can only reappear if it is reintroduced from outside the host-vector community.

The stochastic model we consider is a continuous-time Markov chain. This implies that, given the state of the system at time *t*, the waiting time for the next event (i.e. infections, recovery and death) is exponentially distributed with a rate given by the sum of the rates of all possible events (Bartlett,

1960). The rules for the transition rates are patterned after the corresponding rates in the deterministic model (1)-(2) and are listed in Table 1.

3.1. Branching process in the multi-group model

As well known, pathogen extinction after introduction occurs in stochastic epidemic models either very early or later in time, after at least one major epidemic (Nasell 1999). In order to compute the probability of early extinction, it is possible to linearize the model and analyse the corresponding branching process. Disease extinction and disease invasion in the branching process model correspond to the occurrence of minor and major outbreaks, respectively, in the nonlinear model (Ball 1983).

Branching process models assume that that the numbers of secondary cases caused by each infectious individual are independent and identically distributed (Bartlett 1955; Kendall 1956), implying that the supply of susceptible individuals is not a limiting factor for the outbreak. Then, if the process is also Markov, the number of infectious individuals follows a birth-and-death linear process.

A branching process can then be summarized by the probability generating function of the secondary cases distribution, $G(s)$. When the process is one-dimensional, the probability generating function can be defined as

$$G(s) = \sum_{k=0}^{+\infty} s^k \Pr(Z = k)$$

where $\Pr(Z = k)$ – for $k = 0,1,2,3,\ldots$ – is the probability distribution of the discrete random variable Z, describing the number of new cases caused by each infectious individual (see Harris 1989).

Branching process theory (see Cournot 1847; Athreya & Ney 1972; Haccou et al. 2005) ensures that extinction probability after the introduction of a single infectious individual is the smallest non-negative solution of the equation:

$$G(s) = s\,. \tag{5}$$

In a Markov process, infection transmission can be represented by a Poisson process and infection duration is exponentially distributed, so that secondary cases follow a geometric distribution with mean ν (Diekmann & Heesterbeek 2000) leading to

$$G(s) = \frac{1}{1+\nu(1-s)}.$$

The basic reproduction number is, by definition, the average value of *Z*. Hence, in the case of infection processes $\nu = R_0$. Extending this to multi-type branching processes, as required by the host-vector model, the distributions of secondary infections of each type can be described by an *m*-dimensional generating function (Griffiths 1972; Ball 1983) whose *i*-th component is:

$$G_i(s_1,\ldots,s_m) = \frac{1}{1+\sum_{j=1}^{m} R_{0,i}^{j}(1-s_j)} \qquad (6)$$

where $R_{0,i}^{j}$ represents the average number of secondary type-*i* infections produced by an infected host of type *j*. Extinction probabilities can still be obtained by solving equation (5) which is now *m*-dimensional. A solution $s=(s_1,\ldots,s_m)$ of (5) represents extinction probabilities in the sense that s_i is the probability of extinction after introduction of 1 individual of type *i*.

In the case of vector-borne diseases with one vector and *n* host types, probability generating functions (6) can be written, as already shown in Lloyd et al. (2007), as

$$G_{H_j}(s_{H_1},\ldots,s_{H_n},s_V) = G_{H_j}(s_V) = \frac{1}{1+R_0^{H_jV}(1-s_V)}, \quad j=1,\ldots,n \qquad (7)$$

$$G_V(s_{H_1},\ldots,s_{H_n},s_V) = G_V(s_{H_1},\ldots,s_{H_n}) = \frac{1}{1+\sum_{j=1}^{n}\left[R_0^{VH_j}(1-s_{H_j})\right]}. \qquad (8)$$

Then, by imposing condition (5) in equations (7) and (8), we compute the probability of pathogen extinction following the introduction of a single infected type-*j* host (s_{H_j}) or a single infected vector (s_V) as:

$$G_{H_j}(s_V) = \frac{1}{1 + R_0^{H_j V}(1 - s_V)} = s_{H_j}, \quad j = 1, \ldots, n \tag{9}$$

$$G_V(G_{H_1}(s_V), \ldots, G_{H_n}(s_V)) = \frac{1}{1 + \sum_{j=1}^{n}\left[R_0^{VH_j}\left(1 - \frac{1}{1 + R_0^{H_j V}(1 - s_V)}\right)\right]} = s_V . \tag{10}$$

Equation (10) is obtained by first imposing $G_V = s_V$, then substituting in it s_{Hj} with the LHS of (9). Assuming that π_j is the introduction probability of an infected type-*j* host, it is possible to define the pathogen extinction probability following the introduction of a randomly selected host (s_H) as the weighted sum of the type-*j* extinction probabilities, $\sum_{j=1}^{n} \pi_j s_{H_j}$ (Becker & Marschner 1990; Lloyd et al. 2007). Then

$$\sum_{j=1}^{n} \pi_j s_{H_j} = \sum_{j=1}^{n}\left[\frac{\pi_j}{1 + R_0^{H_j V}(1 - s_V)}\right] = s_H . \tag{11}$$

3.2. Invasion probabilities

System (9)-(10) always have a trivial solution: $s_V = 1,\ s_{H_j} = 1\ \ \forall j = 1, \ldots, n$. In addition, when R_0 defined as in (3) is larger than 1, there also exists a solution with all the *s* strictly lower than 1 (Athreya & Ney 1972); then a major outbreak can occur.

Rewriting expression (10) with $R_0^{VH_j} = \alpha\gamma_j / \delta = r_{VH}\gamma_j$, $R_0^{H_j V} = \beta V \gamma_j / (\zeta H h_j) = r_{HV}\gamma_j / h_j$ as in (3) and rearranging under the assumption that $s_V \neq 1$ (since $s_V = 1$ represents the trivial solution of equation 10), we see that s_V can be obtained as the unique solution in (0,1) of

$$r_{VH} r_{HV} s_V \sum_{j=1}^{n}\left(\frac{\gamma_j^2}{h_j + \gamma_j r_{HV}(1 - s_V)}\right) = 1 . \tag{12}$$

In addition, by assuming as in Becker & Marschner (1990) and Lloyd et al. (2007) that each host individual has the same probability to introduce the pathogen (which corresponds to assume that $\pi_j = h_j$), equation (11) becomes:

$$s_H = \sum_{j=1}^{n} \left(\frac{h_j^2}{h_j + \gamma_j r_{HV}(1 - s_V)} \right). \quad (13)$$

From (12) and (13), it is straightforward to prove that in the case of homogeneous feeding preferences (i.e., $\gamma_j \equiv h_j$) the extinction probabilities ($s_V\big|_{y_j \equiv h_j} = \bar{s}_V$, $s_H\big|_{y_j \equiv h_j} = \bar{s}_H$) in the multi-host branching process are the same as in the one-host case, as shown in Lloyd et al. (2007):

$$\bar{s}_V = \frac{1 + r_{HV}}{r_{HV}(1 + r_{VH})}; \quad \bar{s}_H = \frac{1 + r_{VH}}{r_{VH}(1 + r_{HV})}. \quad (14)$$

Then, if vector-to-host and host-to-vector transmissions are the same ($r_{VH} = r_{HV} = r$), invasion probabilities from an introduced infected vector or random host are the same:

$$1 - \bar{s}_V = 1 - \bar{s}_H = 1 - \frac{1}{r}. \quad (15)$$

Hereafter we show that, in the presence of heterogeneous feeding preferences, the invasion probabilities due to pathogen introduction in the multi-group branching process can significantly differ from the homogeneous case. Specifically, we show that in the presence of heterogeneity in the vector feeding preferences:

i) the invasion probability following the introduction of an infected vector, $p_V = 1 - s_V$, is always larger than in the homogeneous case, i.e. $p_V = 1 - s_V > 1 - \bar{s}_V = \bar{p}_V$ (see Proposition 1 in Appendix A);

ii) on the other hand, the invasion probability following the introduction of a random infected host, $p_H = 1 - s_H$, can be lower than in the homogeneous case despite the increase of $R_{0,\mathrm{rel}}$ (see (4)). In particular, condition $p_H < 1 - \bar{s}_H = \bar{p}_H$ is always satisfied for sufficiently high values of r_{HV}, i.e. $r_{HV} > (1 + \sqrt{1 + r_{VH}})/r_{VH}$. In the specific case of symmetry in the homogeneous components of transmission (i.e. $r_{HV} = r_{VH} = r$) the condition $p_H < \bar{p}_H$ is always satisfied for $r > \varphi$ (where $\varphi = (1 + \sqrt{5})/2$ represents the golden ratio), see Proposition 2 in Appendix A;

iii) moreover, in the case of homogeneity in host group abundance and a large number of host groups ($n \to +\infty$), $p_V = 1 - s_V$ tends to $r_{VH}/(r_{VH}+1)$ and $p_H = 1 - s_H$ tends to zero under the heterogeneous feeding rates that maximize $R_{0,\text{rel}}$ (Propositions 4 and 5 in Appendix A);

In accordance with previous claims, we show that expression (15), which defines the invasion probabilities in the case of homogeneous biting rate, does not hold in the presence of heterogeneous biting rates. Specifically, we show that:

iv) in the case $r_{VH} = r_{HV} = r$, the invasion probability following the introduction of an infected vector is always larger than that of an infected host, i.e. $p_V > p_H$ (see Proposition 3 in Appendix A).

Formal proofs of (i)-(iv) are given in Appendix A.

The previous results show that the invasion probability following the introduction of an infected vector increases with the heterogeneity in hosts-vectors mixing (Propositions 1 in Appendix A) and it can reach its maximum value when $R_{0,\text{rel}}$ is maximized (see Remark on Proposition 4 in Appendix A). On the other hand, the invasion probability following the introduction of an infected host shows a more complex relationship with the heterogeneity in host-vector mixing (Proposition 2 in Appendix A), even showing cases where it can tend to zero when $R_{0,\text{rel}}$ is maximized (see Remark on Proposition 5 in Appendix A). In addition, differently from the one-host model, differences in invasion probabilities via infected vectors (p_V) or hosts (p_H) emerge also in the case of symmetric transmission between vector-to-host and host-to-vector (i.e. $r_{VH} = r_{HV} = r$), see Proposition 3.

3.3 Numerical analyses

Through numerical analyses, we investigate more in-depth the effects of heterogeneity in host spatial, behavioural or genetic structure and vector biting rate on disease establishment. Specifically,

we analyse the role of the number of host types (n), the distribution of host densities (h_j), and the distribution of vector feeding preferences (γ_j) on invasion probabilities.

In order to compare the role of heterogeneous mixing on invasion probabilities via infected vector/host introduction, we assumed in our numerical analyses that the homogeneous components of the vector-to-host (r_{VH}) and host-to-vector (r_{HV}) basic reproduction number are the same, i.e. $r_{VH} = r_{HV} = r$. The previous assumption implies that invasion probabilities p_V and p_H are the same in the presence of a single host type (i.e. n = 1, see Lloyd et al. 2007) or when vectors feed on host types in proportion to their frequency in the community (i.e., $\gamma_j \equiv h_j$, see (15)). As a consequence, the differences observed in our numerical analyses between p_V and p_H are uniquely due to the effect of heterogeneous mixing.

3.3.1 Invasion probabilities and number of host groups

In Fig. 1 we show the effect of the number of host types on invasion probabilities via infected vectors (blue) or hosts (red) for two different values of r (r = 1, panel a; r = 3, panel b), under the assumptions $\gamma_j = \Gamma_j / \sum_j \Gamma_j$ and $h_j = H_j / \sum_j H_j$, where Γ_j and H_j are random extractions from independently distributed standard lognormals, *Lognorm*(0,1). In particular, in Fig. 1 we show invasion probabilities obtained from the numerical solutions of equations (11) and (12) – see filled (dark and shaded) boxes –, together with simulations obtained from the individual-based model as in Tab. 1 – see open dots –. The simulations of the individual-based model were stopped either when no infected individuals were left in the system, or when 10,000 events had occurred. We see a very good agreement between the individual-based model and its branching process approximation. In agreement with Propositions 1 and 2, Fig. 1 shows that p_V always increases with n, while – according to Proposition 2 – p_H increases with n for low values of r (Fig. 1a) and decreases for high values of it (Fig. 1b). In addition, Fig. 1 reveals that asymptotic values of invasion probabilities are approached already for relatively low number of host types (n = 10-20).

[Fig. 1 about here]

3.3.2 Invasion probabilities and the heterogeneity in biting rate

Furthermore, numerical analysis allows to highlight more complex behaviours emerging as a consequence of heterogeneous mixing among hosts and vectors. Fig. 2 displays the effect of the variability in the distributions of biting rates, $\Gamma_j \propto Lognorm(0,\sigma^2)$, and host type frequencies in the community, $H_j \propto Lognorm(0,\sigma^2)$, defined by the log-standard deviation parameter σ of the distributions on invasion probabilities via infected vectors (blue) or hosts (red). Since the heterogeneity in host-vector mixing measured as $\mathrm{var}\left(\gamma_j / h_j; h_j\right)$ increases with σ, we can read Fig. 2 as the effect of increased heterogeneity on the invasion probabilities, ranging from a homogeneous mixing condition, i.e. $\mathrm{var}\left(\gamma_j / h_j; h_j\right) = 0$, when $\sigma = 0$ to a maximum heterogeneity condition when σ is maximized. In particular, Fig. 2 shows that, in the case $r = 1$, the invasion probability p_V, obtained from the solution of equation (12), is always an increasing function of σ with a minimum value $p_V\big|_{\sigma=0} = \bar{p}_V = 1-1/r = 0$ (for $\sigma = 0$), and an asymptotic value for large heterogeneities $p_V\big|_{\sigma\to+\infty} = 1-1/(r+1) = 0.5$ (for large values of σ) as predicted by Proposition 1. On the other hand, p_H, obtained from the solution of equation (11), follows a unimodal pattern with $p_H\big|_{\sigma=0} = p_H\big|_{\sigma\to+\infty} = 0$ and a maximum value for intermediate σ (i.e. heterogeneity).

[Fig. 2 about here]

In Fig. 3 we show a generalisation of the results obtained in Fig. 2. Specifically, we computed the invasion probabilities p_V, in panel A, and p_H, in panel B, as solutions of equations (11) and (12) for different values of r. We show that, similarly to Proposition 4, p_V tends to an asymptotic value $p_V\big|_{\sigma\to+\infty} = 1-1/(r+1)$ for large values of σ, i.e. heterogeneity (see dot-dashed lines in panel A). On the other hand, p_H displays a unimodal pattern for values of r lower than the golden ratio ($r < \varphi = (1+\sqrt{5})/2$), in contrast to a monotonically decreasing function of σ for $r > \varphi$.

[Fig. 3 about here]

3.3.3. Effect of correlation among bites and host groups abundance

In this Section, we present the consequences of removing the hypothesis of independence between biting rate, γ_j, and host type frequency in the community, h_j. In particular, in Fig. 4 we show the invasion probabilities p_V and p_H, obtained by numerically solving equations (11) and (12), as functions of the log-standard deviation parameter σ in biting rates (γ_j) and host type frequencies in the community (h_j), which represents increasing heterogeneities, for different values of Spearman rank correlation between Γ and H log-normal distributions. Correlated lognormal random variables were generated from exponentiating bivariate normal distributions with non-zero cross-covariance. We notice that by increasing the correlation between biting rates and host frequencies, corr(γ_j, h_j), the values of p_V and p_H mutually tend towards each other. However, Fig. 4 highlights that the qualitative behaviour observed in the case of independent distributions holds also in the presence of large value of correlation, e.g. corr(γ_j, h_j) = 0.75.

[Fig. 4 about here]

3.3.4 Alternative assumptions on the introduction probabilities of infected hosts

In Section 3.1. we assumed, as in Becker & Marschner (1990) and Lloyd et al. (2007), that each host individual has the same probability to introduce the pathogen (i.e. $\pi_j = h_j$). Here, we examine the consequences of assuming that the probability of each type-j host to introduce the infection is proportional to the fraction of bites allocated to that type, i.e.

$$\pi_j = \frac{h_j \gamma_j}{\sum_j (h_j \gamma_j)}. \qquad (16)$$

In particular, in Fig. 5 we show the invasion probabilities p_V and p_H, obtained by numerically solving equations (11) and (12) with π_j as in (16), as functions of the log-standard deviation parameter σ in biting rates (γ_j) and host type frequencies in the community (h_j), which represents increasing heterogeneities, for different values of the homogeneous components of the basic

reproduction number ($r_{VH} = r_{HV} = r$) when expression (16) holds. As in Figs. 2 and 3, we find that: (*i*) the invasion probability via infected vector (p_V) is larger than via infected host (p_H), and (*ii*) the invasion probability via infected host (p_H) increases [decreases] as heterogeneity increases for low [high] values of the homogeneous component of the basic reproduction number (r).

[Fig. 5 about here]

4. DISCUSSION AND CONCLUSIONS

In this work, we explored the effect of heterogeneities in host-vector mixing on the probability of invasion of mosquito-borne infections in a multi-group model. In particular, we investigated the role of the distribution of vector feeding preferences and host densities in shaping disease invasion probabilities using a branching process framework. Our analytical and numerical results suggest that increasing the heterogeneity in feeding preferences and/or host densities always increases the invasion probability following the introduction of an infected vector (p_V), see Proposition 1 and Fig. 3a. On the other hand, the invasion probability following the introduction of an infected host (p_H) follows a more complex pattern. Precisely, p_H is a decreasing function of heterogeneity when the homogeneous parts of transmission (r_{HV} and r_{VH}) are high, while a unimodal function when r_{HV} and r_{VH} are low (see Proposition 2 and Fig. 3b). In simple terms, large heterogeneities in host-vector mixing in multi-host models tend to favour the invasion of diseases introduced by vectors and tend to hinder those introduced by hosts. From our results follows the unexpected consequence that, in multi-host models, the invasion probability p_H can exhibit a decreasing relationship with heterogeneity which runs opposite to that of the most classical measure of disease invasion: the basic reproduction number (Dye & Hasibeder 1986); indeed, it can even approach zero at parameter values such that $R_{0,\mathrm{rel}}$ is maximized (see Fig. 3b).

The biological explanation for the observed patterns relies on the remark that, as heterogeneity increases, so does the preference of vectors for certain host types; as a consequence vector bites will be concentrated in a few preferred host groups. In this scenario, highly preferred and little abundant

host types may act as disease super-spreaders. Then, an infected vector is more likely to introduce the infection into a preferred host group from which disease transmission is effective, whereas a randomly introduced infected host is more likely to belong to a host type with ineffective transmission. Thus, introducing an infected vector is much riskier, by the very nature of the heterogeneity.

The complex pattern of invasion probability following the introduction of an infected host is analogous to the effect of individual variations in directly-transmitted infections. In these cases, when transmission is low, heterogeneity initially increases invasion probability, due to the increase in R_0; as heterogeneity is further increased, the effect of heterogeneity itself becomes dominant and causes a decrease in invasion probability. On the other hand, when the homogeneous component of the basic reproduction number is already high, the increase in R_0 due to the heterogeneity becomes less relevant than the increase in the probability to introduce primary hosts with ineffective transmission; this causes p_H to always decrease with heterogeneity.

Differences in the effect of heterogeneities when the disease is introduced by either an infected vector or host have been observed by Smith et al. (2007) by adapting, in the context of a deterministic vector-host model, the basic reproduction number metric to the case of finite host populations. Analogously to our findings, Smith et al. (2007) showed that, in finite host populations, heterogeneous biting amplifies disease transmission in the case of a primary infected vector, while may reduce disease transmission in the case of a primary infected host, especially when the homogeneous component of R_0 is large. The explanation of Smith et al. (2007) findings relies on the remark that, when the host population size is infinite, each infected vector bite lands on a different host. On the other hand, when the host population is low, the same host can receive multiple infected bites, then reducing the basic reproduction number. It is interesting to notice that the biological mechanism shaping the patterns observed by Smith et al. (2007) is completely different from what observed here, where the branching process approximation works under the assumption of infinite populations.

Similarly to the behaviour of p_H found here for high values of the homogeneous component in the transmission terms, several authors showed that, in Susceptible-Infected-Recovered (SIR) model for directly transmitted infections, increasing heterogeneity in transmission among individuals (Lloyd-Smith et al. 2005) or groups (Xiao et al. 2006) leads to a decrease in the disease invasion probability. However, SIR-like models incorporating a metapopulation structure in host distribution found that heterogeneity does not always make extinction more likely. In fact, the hump-shaped behaviour of p_H as a function of heterogeneity (as in Fig. 2) is similar to the relationship observed between infections persistence and variability in between-patch transmission in metapopulations. As shown by Bolker & Grenfell (1995), Keeling (2000), and Hagenaars et al. (2004), the global persistence of infections in SIR-like metapopulation models may be maximized for intermediate levels of spatial heterogeneity. This pattern arises when the coupling among patches is sufficiently strong to generate frequent between-patch transmission (which favours disease re-introduction in different patches), but not so strong that spatial desynchronization (which is needed to avoid global extinctions) is lost. There is definitely an analogy between the mechanism observed in SIR-like metapopulation models and the above described mechanism found here for the invasion via primary infected host.

We also show that the results obtained with the multi-group model significantly differ from those of the homogeneous branching process (thoroughly investigated by Lloyd et al. 2007). Lloyd et al. (2007) proved that the invasion probabilities in the homogeneous model are the same when transmission from host-to-vector and vector-to-host are symmetric (i.e. $r_{HV} = r_{VH}$) following either the introduction of an infected host or vector. Here, we show that, in the case of symmetric transmission, p_V is always larger than p_H (see Proposition 3 and Figs. 2,3,4), suggesting that the introduction of infected mosquitoes can more likely generates major outbreaks than the introduction of infected hosts. Even when host-to-vector transmission is larger than vector-to-host, the multi-group model predicts larger invasion probabilities due to the introduction of primary infected vectors than hosts (i.e. $p_V > p_H$), as long as heterogeneity in feeding preferences and/or host

densities is sufficiently high (see Fig. S1 in the electronic supplementary material); this contrasts with the homogeneous branching process which, in such cases, predicts $p_H > p_V$. This result can provide a general theoretical framework for programs dealing with outbreak control occurring after mosquito-borne pathogen introduction and supporting specific surveillance and control depending on different introduction pathways. For instance, in the case of West Nile virus, it has been shown that the highest risk of virus introduction into the Galapagos Islands is due to infected mosquitoes while the risk due to infected hosts (e.g. avian migration pathway) is much lower (Kilpatrick et al., 2006b). Under our model assumptions, such an introduction would increase the risk of a disease outbreak occurrence, indicating that a greater effort should be devoted to contain infected mosquitoes introduction.

In our numerical analyses we also investigated the role of correlation between biting rate, γ_j, and host type frequency in the community, h_j. Intuitively, we should expect a larger proportion of mosquito bites in more abundant host groups, corresponding to positive correlation values. However, field studies on mosquito feeding preferences performed in avian host communities display a mixed pattern of relationships between biting rate and host abundance. We analysed data published by Hassan et al. (2003, see Tables 2 and 3 therein), Kilpatrick et al. (2006a, see Figure 1 therein), and Thiemann et al. (2011, see columns "Winter" in Table 2 therein), and we found no significant correlations between mosquito blood meals and avian census data, while we found significant positive correlations in data published by Hamer et al. (2009, see Table 2 therein), Spearman-ρ correlation coefficient = 0.55 (p-value = 0.005), and Thiemann et al. (2011, see columns "Late summer" in Table 2 therein), Spearman-ρ = 0.42 (p-value = 0.044). Our numerical analyses show that, even assuming a correlation value higher than those observed in the mentioned studies (i.e. Spearman-ρ = 0.75), results are consistent with those obtained assuming independency between γ_j and h_j (see Fig. 4).

The results we obtained for p_H depend on the assumption that each host individual has the same probability to introduce the pathogen (i.e. $\pi_j = h_j$; as in Becker & Marschner 1990 and Lloyd et al.

2007). While this assumption can be generally accepted when different groups represent spatial or behavioural heterogeneity in the same host population, it can not be adequate in describing disease invasion in groups representing different host species. In effect, different host species may differ in competence levels, as observed in West Nile virus (Kilpatrick et al. 2007) and *Borrelia* infections (Ginsberg et al. 2005). It must be however remarked that, in the case of West Nile virus, Komar et al. (2003) showed that avian host species characterized by different competence display very different levels of infectiousness and duration of viraemia (which affect the ability to spread the disease), but similar levels of susceptibility (which mainly affects the ability to become infected and eventually introduce the pathogen in a new environment through dispersal or migration, i.e. π_j).

We tested the robustness of our results to alternative assumptions on π_j, by computing s_H under the assumption that the probability of each type-*j* host to introduce the infection is proportional to the fraction of bites allocated to him, i.e. see (16). This assumption is equivalent to assume that vector feeding preferences in the community from which the first infected host originated and in the community of destination are similar. When probability of each type-*j* host to introduce the infection is proportional to the fraction of bites allocated to him, we found that, similarly to the case $\pi_j = h_j$, the invasion probability via infected vector is larger that that via infected host, and the invasion probability via infected host increases [decreases] as heterogeneity increases for low [high] values of the homogeneous component of the basic reproduction number (r). However, when r is low, the unimodal pattern of p_H with heterogeneity observed in the case $\pi_j = h_j$, does not hold any longer.

Finally, we remark that our results hold also under different hypotheses than those assumed in numerical simulations on the distribution of blood meals and host group abundance (e,g., power-law distributed, see Fig. S2 in the electronic supplementary material) and the duration of the infectious periods (e.g., fixed instead of exponentially distributed, see Fig. S3 in the electronic supplementary material).

Other ecological and epidemiological aspects that can play a significant role in the invasion of new infections were not included in the present study. Here, we assumed a constant and homogeneous vector population. However, mosquito species can display marked seasonal dynamics with population abundance depending on several biotic and abiotic factors (Kilpatrick & Pape 2013, Rosà et al. 2014) and spatial heterogeneity (Smith et al. 2004); moreover, infections can be transmitted by multiple vectors (Goddard et al. 2002). These aspects are likely to affect significantly the probability of disease invasion and the occurrence of major outbreaks. In particular, in the case of spatial heterogeneity, the assumption of homogenous vector population coupled with a patchy host population may not be acceptable. However, we expect our results to hold also when the assumption of homogenous vector population is violated, if the vector feeding preferences among patches are similar (analogously to what shown in Hasibeder & Dye 1988 for the basic reproduction number). On the other hand, when feeding preferences significantly change among patches, we expect more complex behaviours to emerge, especially for infections introduced via primary infected vectors. Despite the limitations acknowledged above, we are confident that our work, thanks to the simplicity and generality of the used framework, will serve as a useful baseline for future investigations.

APPENDIX A. Computation of invasion probabilities

Proposition 1. *The invasion probability following the introduction of an infected vector,* $p_V = 1 - s_V$, *defined as the solution of (12), satisfies the inequalities*

$$p_V > \bar{p}_V = 1 - \bar{s}_V = 1 - \frac{r_{HV} + 1}{r_{HV}(r_{VH} + 1)} \tag{A.1}$$

and

$$p_V < \frac{r_{VH}}{r_{VH} + 1} \tag{A.2}$$

Inequality (A.1) states that invasion probability with heterogeneity is higher than in the homogeneous case (compare (14)), while (A.2) gives an absolute upper bound. In order to prove Proposition 1, we start from two simple Lemmas. In both, for ease of notation, we use the functions

$$f_j(s) = h_j + \gamma_j r_{HV}(1 - s).$$

Lemma 1.

$$\sum_{j=1}^{n} \frac{\gamma_j^2}{f_j(s)} \geq \frac{1}{1 + r_{HV}(1 - s)} \tag{A.3}$$

with equality if and only if $\gamma_j = h_j$ *for all* $j = 1,\ldots,n$.

Proof.

$$\begin{aligned} 0 \leq \sum_{j=1}^{n} \left(\frac{\gamma_j}{f_j(s)} - \frac{1}{1 + r_{HV}(1 - s)} \right)^2 f_j(s) = \ldots \\ \ldots = \sum_{j=1}^{n} \frac{\gamma_j^2}{f_j(s)} - \frac{2}{1 + r_{HV}(1 - s)} + \frac{1 + r_{HV}(1 - s)}{(1 + r_{HV}(1 - s))^2} = \sum_{j=1}^{n} \frac{\gamma_j^2}{f_j(s)} - \frac{1}{1 + r_{HV}(1 - s)}. \end{aligned} \tag{A.4}$$

Note that, in order to obtain the second line in (A.4), we used the identities

$$\sum_{j=1}^{n} \gamma_j = 1 \text{ and } \sum_{j=1}^{n} f_j(s) = 1 + r_{HV}(1 - s).$$

(A.4) yields the inequality (A.3). Equality in (A.4) holds if and only if

$$\frac{\gamma_j}{f_j(s)} = \frac{1}{1+r_{HV}(1-s)} \;\forall\; j=1,\ldots,n \quad \Leftrightarrow \quad \gamma_j = h_j \;\forall\; j=1,\ldots,n$$

yielding the thesis. □

Lemma 2.

$$\sum_{j=1}^{n} \frac{\gamma_j^2}{f_j(s)} < \frac{1}{r_{HV}(1-s)}. \tag{A.5}$$

Proof. We take $\sum_{j=1}^{n}\left(\gamma_j^2 / f_j(s)\right)$ as a function of $\gamma_1,\ldots,\gamma_n$:

$$G(\gamma_1,\ldots,\gamma_n) = \sum_{j=1}^{n} g_j(\gamma_j) \text{ with } g_j(x) = \frac{x^2}{h_j + x r_{HV}(1-s)}.$$

It is easy to see, by carefully computing its first and second derivatives, that, for any $s \le 1$, each $g_j(x)$ is an increasing and convex function of $x \ge 0$.

Then G is a convex function, so that its maximum on the simplex $\{\gamma_1,\ldots,\gamma_n : \gamma_i \ge 0, \gamma_1 + \ldots + \gamma_n = 1\}$ will be obtained on one corner.

Assuming, without loss of generality, $0 \le h_1 \le h_2 \le \ldots \le h_n$, it is easy to see that, for each $(\gamma_1, \gamma_2, \ldots, \gamma_n)$ in the simplex, it holds

$$G(\gamma_1,\gamma_2,\ldots,\gamma_n) \le G(1,0,\ldots,0) = \frac{1}{h_1 + r_{HV}(1-s)} \le \frac{1}{r_{HV}(1-s)}. \tag{A.6}$$

□

The strict inequality in (A.6) arises from the fact that all populations are assumed to have positive densities. More generally, note that the maximum $1/(1+r_{HV}(1-s))$ will be approached as preference γ_j approaches 1 on a species whose density approaches 0, something not very biologically realistic.

Proof of Proposition 1. As noted above s_V can be obtained as the solution of (12) that we can rewrite as

$$H(s) := r_{HV} r_{VH} s \sum_{j=1}^{n} \frac{\gamma_j^2}{f_j(s)} = 1 \qquad \text{(A.7)}$$

It is immediate to see that H is an increasing function, $H(0) = 0$ and $H(1) = r_{HV} r_{VH} \sum_{j=1}^{n} \gamma_j^2 / h_j = R_0$.

Hence, when $R_0 > 1$, (A.7) has a unique solution in (0,1) which is the extinction probability s_V.

Now Lemmas 1 and 2 yield immediately

$$H_1(s) \le H(s) < H_2(s) \text{ with } H_1(s) = \frac{r_{HV} r_{VH} s}{1 + r_{HV}(1-s)} \text{ and } H_2(s) = \frac{r_{VH} s}{1-s}. \qquad \text{(A.8)}$$

Note that H_1 and H_2 are also increasing functions satisfying $H_1(0) = H_2(0) = 0$, so that there exists unique $s_1^* > 0$ and $s_2^* > 0$ such that $H_1(s_1^*) = H_2(s_2^*) = 1$. These equations can be easily solved, yielding $s_2^* = \frac{1}{1 + r_{VH}}$ and $s_1^* = \frac{1 + r_{HV}}{r_{HV}(1 + r_{VH})}$.

The inequalities (A.8) imply $s_2^* < s < s_1^*$, which are (A.1) and (A.2). □

Remark 1. Note that, using (A.4) in Lemma 1, $H(s)$ can be written as

$$H(s) = r_{HV} r_{VH} s \left[\frac{1}{1 + r_{HV}(1-s)} + \mathrm{var}\left(\frac{\gamma_j}{f_j(s)}; f_j(s) \right) \right].$$

Following the same argument as in the proof of Proposition 1, it follows that, if heterogeneity $\mathrm{var}(\gamma_j / h_j; h_j)$ is increased in the sense of increasing $\mathrm{var}(\gamma_j / f_j(s); f_j(s))$, necessarily s_V decreases, hence the probability of invasion increases.

Unfortunately, we were not able to find a simple relation between the heterogeneity $\mathrm{var}(\gamma_j / h_j; h_j)$ and $\mathrm{var}(\gamma_j / f_j(s); f_j(s))$; however, it seems quite plausible that if γ_j and h_j depend (as in the numerical analyses) on a single parameter σ such that $\mathrm{var}(\gamma_j / h_j; h_j)$ increases with σ, the same will be true for $\mathrm{var}(\gamma_j / f_j(s); f_j(s))$, so that s_V will decrease with σ. In this sense, we believe we can safely state that s_V decreases with heterogeneity.

We now switch to considering p_H, the probability of infection invasion when a random host is introduced in the population. We start from (13) that can be rewritten as

$$s_H = \sum_{j=1}^{n} \frac{h_j^2}{f_j(s_V)} = \sum_{j=1}^{n} \left[\frac{h_j^2}{f_j(s_V)} - \frac{r_{HV}^2 (1-s_V)^2 \gamma_j^2}{f_j(s_V)} + \frac{r_{HV}^2 (1-s_V)^2 \gamma_j^2}{f_j(s_V)} \right]$$

Simple algebraic manipulations, remembering that $f_j(s) = h_j + \gamma_j r_{HV}(1-s)$, then yields

$$\begin{aligned} s_H &= \sum_{j=1}^{n} \frac{\left(h_j - r_{HV}(1-s_V)\gamma_j\right)\left(h_j + r_{HV}(1-s_V)\gamma_j\right)}{f_j(s_V)} + r_{HV}^2 (1-s_V)^2 \sum_{j=1}^{n} \frac{\gamma_j^2}{f_j(s_V)} = \dots \\ \dots &= 1 - r_{HV}(1-s_V) + \frac{r_{HV}(1-s_V)^2}{r_{VH} s_V} H(s_V) = 1 - r_{HV}(1-s_V) + \frac{r_{HV}(1-s_V)^2}{r_{VH} s_V} \end{aligned} \qquad \text{(A.9)}$$

finally using $H(s_V) = 1$.

Equation (A.9) can be summarised as $s_H = q(s_V)$ where

$$q(s_V) = 1 - r_{HV}(1-s_V) + \frac{r_{HV}(1-s_V)^2}{r_{VH} s_V}.$$

Proposition 2. *(1) If* $r_{HV} \geq \dfrac{1+\sqrt{1+r_{VH}}}{r_{VH}}$, $p_H \leq \bar{p}_H$, *the value for the homogeneous case.*

Furthermore, if p_V *increases with a parameter* σ *describing heterogeneity,* p_H *decreases with* σ.

(2) If $r_{HV} < \dfrac{1+\sqrt{1+r_{VH}}}{r_{VH}}$, p_H *can becomes larger than* $\bar{p}_H$. *More precisely, if* p_V *increases with a heterogeneity parameter* σ, p_H *initially increases up to a maximum, then decreases.*

Proof. We always work with the quantities $s_H = 1 - p_H$ and $s_V = 1 - p_V$, as they appear in equations (A.7) and (A.9).

First of all, observe that $q(.)$ is a decreasing function in $(0, s_m)$ and increasing in $(s_m, 1)$ where

$$s_m = \frac{1}{\sqrt{1+r_{VH}}}.$$

Then, if $\bar{s}_V \leq s_m$, decreasing s_V as a result of heterogeneity will increase $s_H = q(s_V)$. A simple

calculation show that this happens if and only if $r_{HV} \geq \frac{1+\sqrt{1+r_{VH}}}{r_{VH}}$.

Vice versa, if $\bar{s}_V > s_m$, decreasing s_V results in a decrease of s_H up to the minimum $q(s_m)$.

Decreasing s_V further will result then in an increase of s_H. □

The previous Proposition states that the invasion probability following the introduction of a random infected host will often be lower than in the homogeneous case despite the increase of $R_{0,\text{rel}}$ (see (4)). In particular, it will always be lower for sufficiently high values of r_{HV}, i.e. $r_{HV} \geq \frac{1+\sqrt{1+r_{VH}}}{r_{VH}}$.

In the symmetric case $r_{VH} = r_{HV} = r$, the condition $p_H \leq \bar{p}_H$ is always satisfied for $r > \varphi$, the golden ratio.

For the symmetric case, the identity (15) states that, in the case of homogeneous biting rates, the invasion probability is the same after the introduction of a random infected host or an infected vector. We now show that this is not true in the presence of heterogeneous biting rates. Specifically, we show:

Proposition 3. *Given $r_{VH} = r_{HV} = r$, the invasion probability following the introduction of a random infected host in the presence of heterogeneity in vector feeding preferences, p_H, can not be larger than that for an infected vector, i.e. p_V.*

Proof. We need to show $s_H = q(s_V) > s_V$. Under the assumption $r_{VH} = r_{HV} = r$,

$$q(s) > s \quad \Leftrightarrow \quad 1 - r(1-s) + \frac{(1-s)^2}{s} > s \quad \Leftrightarrow \quad rs^2 - s(1+r) + 1 > 0 \quad \Leftrightarrow \quad s < 1/r$$

considering only $s \in (0,1)$.

Now, Proposition 1 states, for the case $r_{VH} = r_{HV} = r$, exactly $s_V < 1/r$. Hence, $q(s_V) > s_V$. □

In case each host type in (1)-(2) contains the same number of individuals – i.e. $h_j = 1/n \; \forall j = 1,\ldots,n$ –, Dye and Hasibeder (1986) proved that the relative reproduction number in (4) can be written as $R_{0,rel} = 1 + n^2 \, \mathrm{var}(\gamma_j; 1/n)$, which is an always increasing function of the variance in mosquito feeding preferences. Under the same conditions, we will compute the asymptotic behaviour of the invasion probabilities p_V and p_H when the variance in mosquito feeding preferences (and consequently $R_{0,rel}$) tends to maximize its value.

As shown by Dye and Hasibeder (1986),

$$\mathrm{var}\left(\gamma_j; \frac{1}{n}\right) = \frac{1}{n}\sum_{j=1}^{n} \gamma_j^2 - \frac{1}{n^2}. \qquad \text{(A.10)}$$

Since

$$\sum_{j=1}^{n} \gamma_j^2 = 1 - \sum_{j=1}^{n} \sum_{i \neq j} \gamma_j \gamma_i ,$$

it follows that (A.10) is maximized when there exists a $k \in [1,\ldots,n]$ such that $\gamma_k \to 1$ (and all others $\gamma_{j \neq k} \to 0$). Then, when $\mathrm{var}(\gamma_j; 1/n) \to \max[\mathrm{var}(\gamma_j; 1/n)] = 1/n - 1/n^2$ the following propositions hold:

Proposition 4. *Let us assume there exists a* $k \in [1,\ldots,n]$ *such that* $\gamma_k \to 1$, $\gamma_{j \neq k} \to 0$ *and that* $h_j = 1/n \; \forall j = 1,\ldots,n$ *; then, the invasion probability following the introduction of an infected vector,* $p_V = \hat{p}_V$*, as defined in (12) can be written as:*

$$\hat{p}_V = 1 - \frac{r_{HV} + 1/n}{r_{HV}(r_{VH} + 1)}. \qquad \text{(A.11)}$$

Proof. From the assumptions follows that

$$\sum_{j=1}^{n} \frac{\gamma_j^2}{h_j + \gamma_j r_{HV}(1 - s_V)} = \frac{1}{1/n + r_{HV}(1 - \hat{s}_V)}. \qquad \text{(A.12)}$$

with $\hat{s}_V = 1 - \hat{p}_V$.

By substituting (A.12) in (12) and rearranging, we obtained the following expression for $\hat{s}_V$

$$\hat{s}_V = \frac{r_{HV} + 1/n}{r_{HV}(r_{VH} + 1)}, \tag{A.13}$$

which proves the proposition. □

Remark 2. For very large numbers of host type populations (i.e., $n \to +\infty$) the invasion probability following the introduction of an infected vector ($\hat{p}_V$), as defined in (A.11), tends towards the maximum value of p_V as defined in proposition 2, i.e.:

$$\hat{p}_V\big|_{n\to+\infty} = \max(p_V) = \frac{r_{VH}}{r_{VH} + 1}.$$

Proposition 5. *Let us assume there exists a* $k \in [1,\ldots,n]$ *such that* $\gamma_k \to 1$, $\gamma_{j\neq k} \to 0$ *and that* $h_j = 1/n \;\forall j = 1,\ldots,n$ *; then, the invasion probability following the introduction of an infected host,* $p_H = \hat{p}_H$ *, as defined in (13) can be written as:*

$$\hat{p}_H = 1 - \frac{r_{VH} + 1}{r_{VH} n(r_{HV} n + 1)} - \frac{(n-1)}{n} \tag{A.14}$$

Proof. From the assumptions follows that expression (13) can be written as

$$\hat{s}_H = \frac{1}{n + n^2 r_{HV}(1 - \hat{s}_V)} + \frac{(n-1)}{n}, \tag{A.15}$$

with $\hat{s}_H = 1 - \hat{p}_H$. By substituting $\hat{s}_V$ as defined in (A.13), expression (A.14) becomes

$$\hat{s}_H = \frac{r_{VH} + 1}{r_{VH} n(r_{HV} n + 1)} + \frac{(n-1)}{n},$$

which proves the proposition. □

Remark 3. For very large numbers of host type populations (i.e., $n \to +\infty$) the invasion probability following the introduction of an infected host ($\hat{p}_H$), as defined in (A.14), tends towards zero ($\hat{p}_H\big|_{n\to+\infty} = 0$).

ACKNOWLEDGEMENTS

The work was funded by EU grant FP7 – 261391 EuroWestNile and partially by the Autonomous Province of Trento (Italy), Research funds for Grandi Progetti, Project LExEM (Laboratory of Excellence for Epidemiology and modelling).

REFERENCES

Adams, B., Holmes, E. C., Zhang, C., Mammen, M. P. Jr, Nimmannitya, S., Kalayanarooj, S. & Boots, M. 2006 Cross-protective immunity can account for the alternating epidemic pattern of dengue virus serotypes circulating in Bangkok. *Proc. Natl Acad. Sci. USA* **103**, 14234-14239.

Andersson, H. & Britton, T. 2000 Stochastic epidemic models and their statistical analysis. Springer Lecture Notes in Statistics. Springer-Verlag, New York.

Aron, J. L. 1988 Mathematical modeling of immunity to malaria. *Math. Biosci.* **90**, 385-396.

Athreya, K. B. & Ney, P. E. 1972 *Branching processes.* Springer, Berlin, Germany.

Bailey, N. T. J. 1975. The Mathematical Theory of Infectious Diseases and its Applications. Griffin, London.

Ball, F. 1983 The threshold behaviour of epidemic models. *J. Appl. Prob.* **20**, 227-241.

Barbour, A. D. 1978 Macdonald's model and the transmission of bilharzia. *T. Roy. Soc. Trop. Med.* H. **72**, 6-15.

Bartlett, M. S. 1955 *An introduction to stochastic processes*. Cambridge University Press, London, UK,.

Bartlett, M. S. 1960 *Stochastic population models in ecology and epidemiology*. Methuen, London, UK.

Bartlett, M. S. 1964 The relevance of stochastic models for large-scale epidemiological phenomena. *Appl. Stat.* **13**, 2-8.

Becker, N. & Marschner, I. 1990 The effect of heterogeneity on the spread of disease. In *Stochastic processes in epidemic theory*, vol. 86 (eds J.-P. Gabriel, C. Lefèvre & P. Picard). Lecture notes in biomathematics, pp. 90–103. Springer, Berlin, Germany.

Bolker, B. & Grenfell, B. 1995 Space, persistence and dynamics of measles epidemics. *Phil. Trans. R. Soc. B* **348**, 309-320.

Cournot, A.-A. 1847 *De l'origine et des limites de la correspondance entre l'algèbre et la géométrie*. Hachette, Paris, France.

Diekmann, O., Heesterbeek, J., & Metz, 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.* 365–382.

Diekmann, O. & Heesterbeek, J. A. P. 2000 *Mathematical epidemiology of infectious diseases*. Wiley, Chichester, UK.

Dietz, K. 1971 Malaria models. *Adv. Appl. Probability* **3**, 208–210.

Dye, C. M. & Hasibeder, G. 1986 Population dynamics of mosquito-borne disease: effects of flies which bite some people more frequently than others. *Trans. Roy. Soc. Trop. Med. Hyg.* **83**, 69-77.

Ferguson, H. M. & Read, A. F. 2004 Mosquito appetite for blood is stimulated by *Plasmodium chabaudi* infections in themselves and their vertebrate hosts. *Malaria J.* **3**, 12.

Ferguson, N., Anderson, R., & Gupta, S. 1999 The effect of antibody-dependent enhancement on the transmission dynamics and persistence of multiple strain pathogens. *Proc. Natl Acad. Sci. USA* **96**, 790-794.

Ginsberg, H. S., Buckley, P. A., Balmforth, M. G., Zhioua, E., Mitra, S. & Buckley, F. G. 2005 Reservoir competence of native north American birds for the Lyme disease spirochete, *Borrelia burgdorferi*. *J. Med. Entomol.* **42**, 445-449.

Goddard, L. B., Roth, A. E., Reisen, W. K. & Scott, T. W. 2002 Vector competence of California mosquitoes for West Nile virus. *Emerging Infectious Diseases* **8**, 1385-1391.

Griffiths, D. A. 1972 A bivariate birth-death process which approximates to the spread of a disease involving a vector. *J. Appl. Prob.* **9**, 65-75.

Hagenaars, T. J., Donnelly, C. A., & Ferguson, N. M. 2004 Spatial heterogeneity and the persistence of infectious diseases. *J. Theor. Biol.* 229, 349-359.

Hamer, G. L., Kitron, U. D., Goldberg, T. L., Brawn, J. D., Loss, S. R., Ruiz, M. O., Hayes, D. B., & Walker, E. D. 2009 Host selection by *Culex pipiens* mosquitoes and West Nile virus amplification. *Am. J. Trop. Med. Hyg.* 80, 268-278.

Harris, T. E. 1989 The Theory of Branching Processes. Dover, New York, USA.

Hasibeder, G., Dye, C. 1988 Population dynamics of mosquito-borne disease: persistence in a completely heterogeneous environment. *Theor. Popul. Biol.* **33**, 31-53.

Hassan, H. K., Cupp, E. W., Hill, G. E., Katholi, C. R., Klingler, K., & Unnasch, T. R. 2003 Avian host preference by vectors of eastern equine encephalomyelitis virus. *Am. J. Trop. Med. Hyg.* 69, 641-647.

Haccou, P., Jagers, P., & Vatutin,V. A. 2005 Branching Processes: Variation, Growth and Extinction of Populations. Cambridge University Press.

Keeling, M. J. 2000 Metapopulation moments: coupling, stochasticity and persistence. *J. Anim. Ecol.* **69**, 725-736.

Kendall, D. G. 1956 Deterministic and stochastic epidemics in closed populations. *Proc. 3rd Berkeley Symp. Math. Statist. Prob.* **4**, 149-165.

Kilpatrick, A. M., Daszak, P., Jones, M. J., Marra, P. P. & Kramer, L. D. 2006a Host heterogeneity dominates West Nile virus transmission. *Proc. R. Soc. B* **273**, 2327-2333.

Kilpatrick, A. M., Daszak, P., Goodman, S. J., Rogg, H., Kramer, L. D., Cedeño, V. & Cunningham, A. A. 2006b Predicting pathogen introduction: West Nile virus spread to Galápagos. *Conserv. Biol.* **20**, 1224-31.

Kilpatrick, A. M., LaDeau, S. L. & Marra, P. P. 2007 Ecology of West Nile virus transmission and its impact on birds in the Western Hemisphere. *Auk* **124**, 1121-1136.

Kilpatrick, A. M., Pape, W. J. 2013 Predicting human West Nile virus infections with mosquito surveillance data. Am J Epidemiol **178**, 829-835.

Komar, N., Langevin, S., Hinten, S., Nemeth, N., Edwards, E., Hettler, D., Davis, B., Bowen, R. & Bunning,M. 2003 Experimental infection of north American birds with the New York 1999 strain of West Nile virus. *Emerg. Infect. Dis.* **9**, 311-322.

Lloyd, A. L., Zhang, J. & Root, A. M. 2007 Stochasticity and heterogeneity in host–vector models. *J. R. Soc. Interface* **4**, 851-863.

Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E. & Getz, W. M. 2005 Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355-359.

Macdonald, G. 1952 The analysis of equilibrium in malaria. *Trop. Dis. Bull.* **49**, 813–829.

Magori, K., Legros, M., Puente, M. E., et al. 2009 Skeeter buster: a stochastic, spatially explicit modelling tool for studying *Aedes aegypti* population replacement and population suppression strategies. *PLoS Neglect. Trop. Dis.* **3**, e508.

Mollison, D. 1991 Dependence of epidemic and population velocities on basic parameters. *Math. Biosci.* **107**, 255-287.

Nasell, I. 1999 On the time to extinction in recurrent epidemics. *J. R. Statist. Soc.* **61**, 309-330.

Perkins, T. A., Scott, T. W., Le Menach, A., & Smith, D. L. 2013 Heterogeneity, mixing, and the spatial scales of mosquito-borne pathogen transmission. *PLoS Comput. Biol.* **7**, e1002174.

Reiner, R. C. Jr, Perkins, T. A. *et al.* 2013 systematic review of mathematical models of mosquito-borne pathogen transmission: 1970–2010. *J. R. Soc. Interface* **10**, 20120921.

Roiz, D., Vazquez, A., Rosà, R., Munoz, J., Arnoldi, D., et al. 2012 Blood meal analysis, flavivirus screening, and influence of meteorological variables on the dynamics of potential mosquito vectors of West Nile virus in northern Italy. *Journal of Vector Ecology* **37**, 1-9.

Rosà, R., Marini, G., Bolzoni, L., Neteler, M., Metz, M., et al. 2014 Early warning of West Nile virus mosquito vector: climate and land use models successfully explain phenology and abundance of *Culex pipiens* mosquitoes in north-western Italy. Parasites & Vectors **7**, 269.

Ross, R. 1911 The prevention of malaria. 2nd edn. J. Murray, London, UK.

Shirai, Y., Funada, H., Seki, T., Morohashi, M. & Kamimura, K. 2004 Landing preference of *Aedes albopictus* (Diptera: Culicidae) on human skin among ABO blood groups, secretors or nonsecretors, and ABH antigens. *J. Med. Entomol.* **41**, 796-799.

Smith, D. L., Dushoff, J. & McKenzie 2004 the risk of a mosquito-borne infection in a heterogeneous environment. *PLoS Biol.* **2**, e368.

Smith, D. L., McKenzie, F. E., Snow, R. W. & Hay, S. I. 2007 Revisiting the basic reproductive number for malaria and its implications for malaria control. *PLoS Biol.* **5**, e42.

Smith, D. L., Perkins, T. A., Reiner, R. C. Jr, Barker, C. M., Niu, T., et al. 2014 Recasting the theory of mosquito-borne pathogen transmission dynamics and control. *Trans. R. Soc. Trop. Med. Hyg.* 108, 185-197.

Thiemann, T. C., Wheeler, S. S., Barker, C. M., & Reisen, W. K. 2011 Mosquito Host Selection Varies Seasonally with Host Availability and Mosquito Density. *PLoS Negl. Trop. Dis.* **5**, e1452.

Verhulst, N. O., Qiu, Y. T., Beijleveld, H. *et al.* 2011 Composition of human skin microbiota affects attractiveness of malaria mosquitoes. *PLoS ONE* **6**, e28991.

Woolhouse, M. E. J., Dye, C., Etard, J.-F., Smith, T., Charlwood, J. D., Garnett, G. P., Hagan, P., Hii, J. L. K., Ndhlovu, P. D., Quinnell, R. J., Watts, C. H., Chandiwana, S. K. & Anderson, R. M. 1997 Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proc. Natl Acad. Sci. USA* **94**, 338-342.

Xiao, Y., Clancy, D., French, N. P. & Bowers, R. G. 2006 A semi-stochastic model for Salmonella infection in a multi-group herd. *Math. Biosci.* **200**, 214-233.

Zainulabeuddin, S. & Leal, W. S. 2007 Maxillary palps are broad spectrum odorant detectors in *Culex quinquefasciatus*. *Chem. Senses* **32**, 727-738.

TABLES

Event	Transition rule	Rate
infection of host *j*	$Y_j \to Y_j + 1$	$\alpha\gamma_j I(H_j - Y_j)/H_j$
infection of vector	$I \to I + 1$	$\beta(V - I)\sum_{j=1}^{n}(\gamma_j Y_j / H_j)$
recovery of host *j*	$Y_j \to Y_j - 1$	ζY_j
death of vector	$I \to I - 1$	δI

Table 1: Rules for the transition rates in the stochastic model

FIGURE CAPTIONS

Figure 1: The invasion probabilities p_V (in blue) and p_H (in red) as functions of the number of host types, n, in the presence of heterogeneous biting rate (γ_j) and host type frequency in the community (h_j). Filled squares represent the median values and shaded boxes represent the inter-quartile ranges of the invasion probabilities obtained by solving equations (11) and (12); open dots represent the invasion probabilities obtained from 10,000 runs of the individual-based model (see Tab. 1). Parameters: $\Gamma, H \propto Lognorm(0,1)$; $\gamma_j = \Gamma_j / \sum_j \Gamma_j$ and $h_j = H_j / \sum_j H_j$; $\pi_j = h_j$; $r_{VH} = r_{HV} = 1$ (panel A); $r_{VH} = r_{HV} = \sqrt{3}$ (panel B). For the individual-based model: $\delta = \xi = 1$; $H = V = 10{,}0000$. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this article.)

Figure 2: The invasion probabilities p_V (in blue) and p_H (in red) obtained by solving equations (11) and (12) as functions of the log-standard deviation parameter σ in biting rates (γ_j) and host type frequencies in the community (h_j), which represents increasing heterogeneities. Tick lines represent the median values, while shaded areas represent the inter-quartile range. Other parameters: $r_{VH} = r_{HV} = 1$; $n = 50$; $\Gamma, H \propto Lognorm(0, \sigma^2)$; $\gamma_j = \Gamma_j / \sum_j \Gamma_j$; and $h_j = H_j / \sum_j H_j$; $\pi_j = h_j$. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this article.)

Figure 3: The invasion probabilities p_V (panel A) and p_H (panel B) obtained by solving equations (11) and (12) as functions of the log-standard deviation parameter σ in biting rates (γ_j) and host type frequencies in the community (h_j), which represents increasing heterogeneities, for different $r_{VH} = r_{HV} = r$. Tick lines represent the median values, while shaded areas represent the inter-quartile range. Dot-dashed lines represent the asymptotic value of p_V as in Proposition 1 in

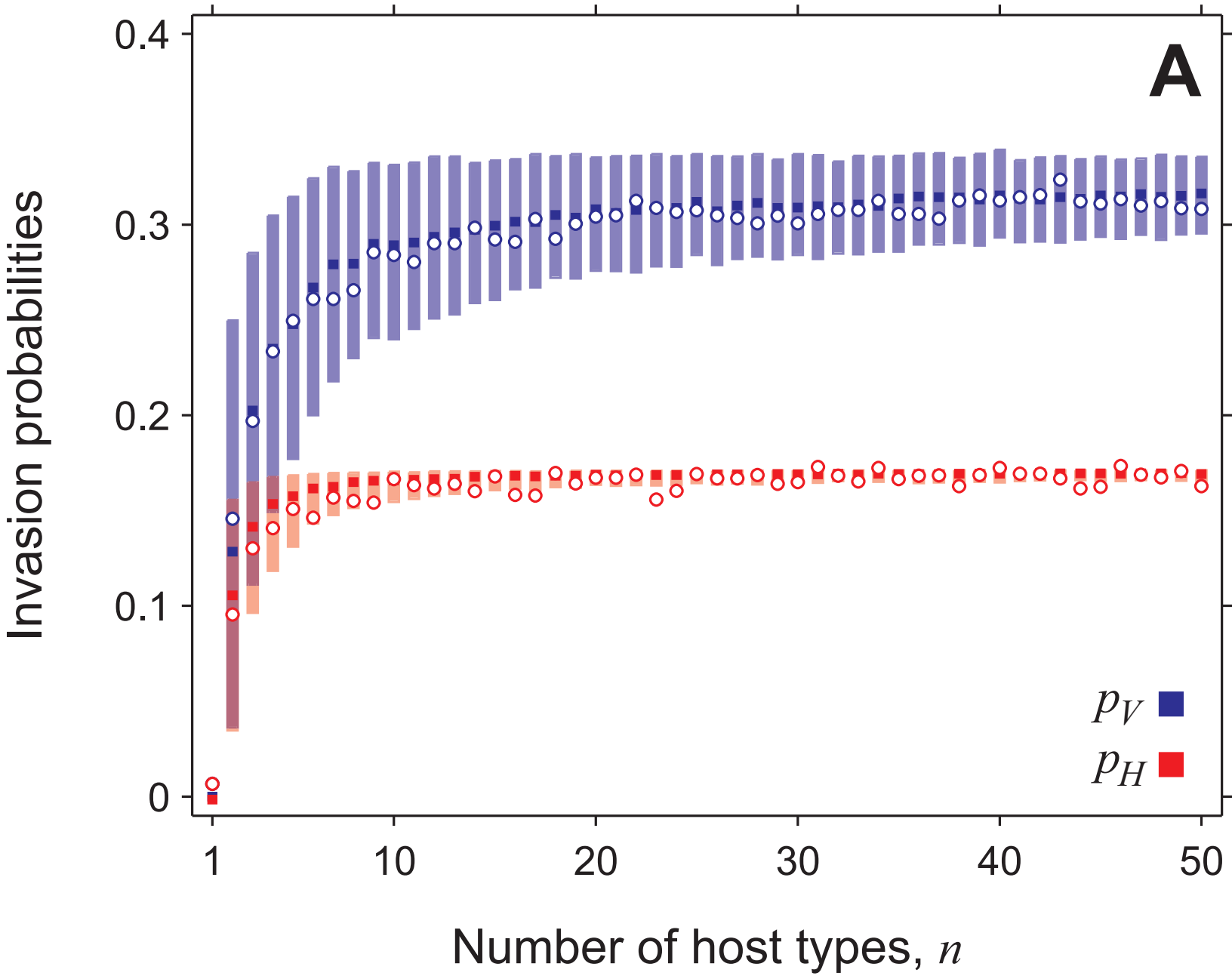
Appendix A. $r = 0.75, 1, \varphi, 3$ (where φ represents the golden number). Other parameters as in Fig. 2. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this article.)

Figure 4: Median values of invasion probabilities p_V (in blue) and p_H (in red) obtained by solving equations (11) and (12) as functions of the log-standard deviation parameter σ in biting rates (γ_j) and host type frequencies in the community (h_j), which represents increasing heterogeneities, for different values of spearman rank correlation between γ_j and h_j. Solid lines: corr(γ_j, h_j) = 0.25; dashed lines: corr(γ_j, h_j) = 0.5; dot-dashed lines: corr(γ_j, h_j) = 0.75. Other parameters as in Fig. 2. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this article.)

Figure 5: Median values of invasion probabilities p_V (in blue) and p_H (in red) obtained by solving equations (11) and (12) as functions of the log-standard deviation parameter σ in biting rates (γ_j) and host type frequencies in the community (h_j), which represents increasing heterogeneities, for two different values of r = 1 (solid lines) and r = 3 (dashed lines), and under the assumption $\pi_j = h_j\gamma_j / \sum_j (h_j\gamma_j)$; where π_j represents the introduction probability of an infected type-j host. Other parameters as in Fig. 2. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this article.)

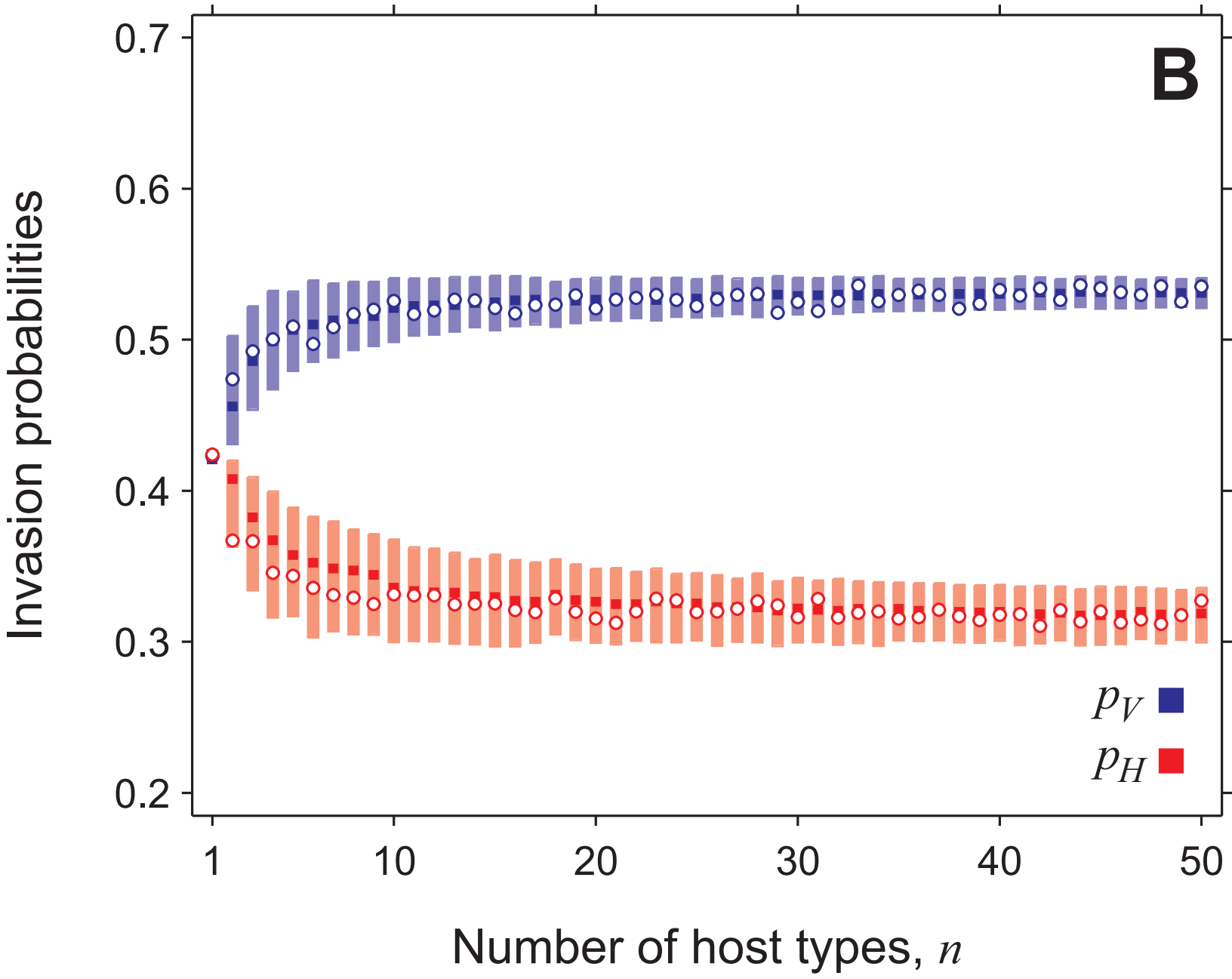
4. Figure 1a

Click here to download 4. Figure: fig1a.eps



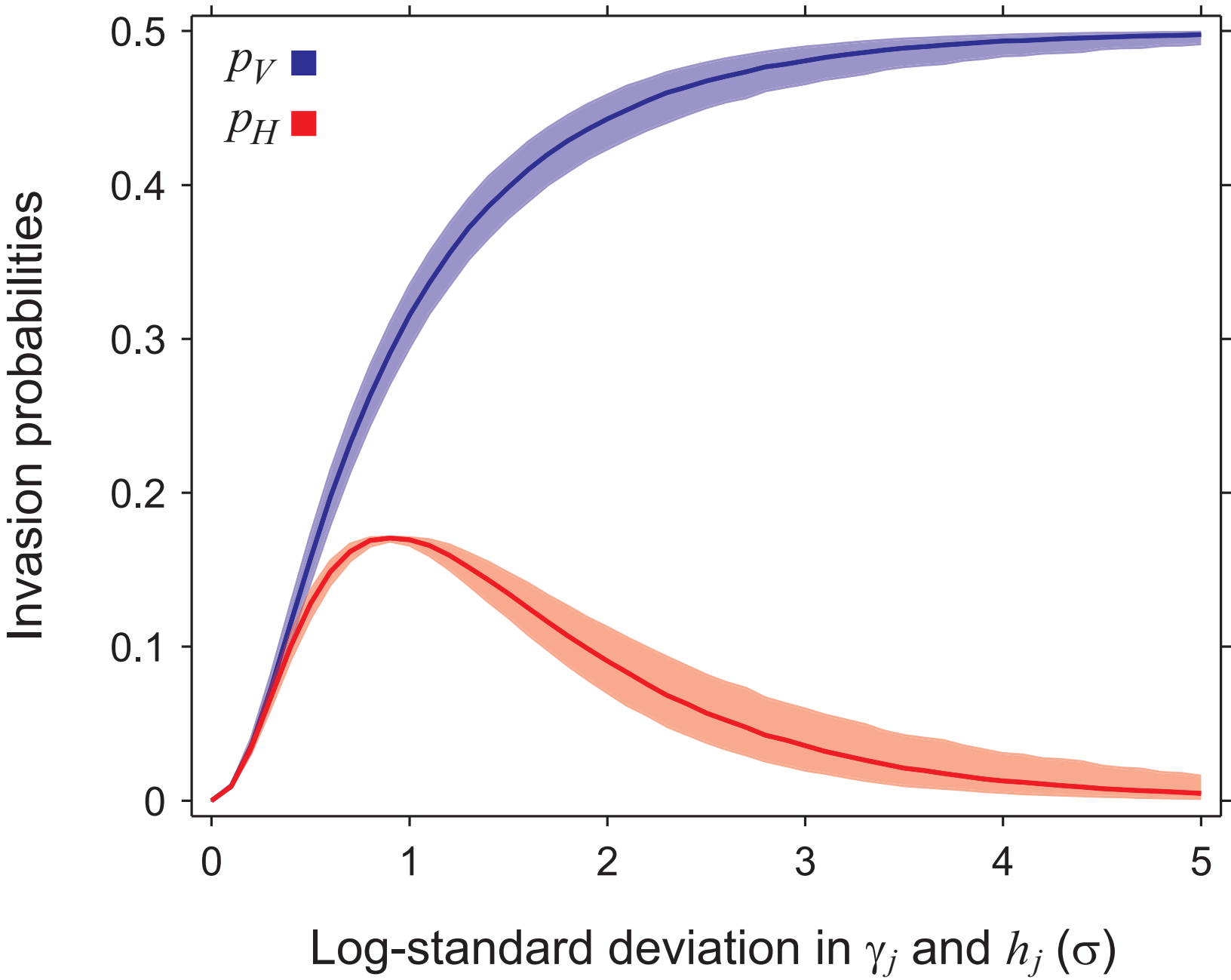
4. Figure 1b

Click here to download 4. Figure: fig1b.eps



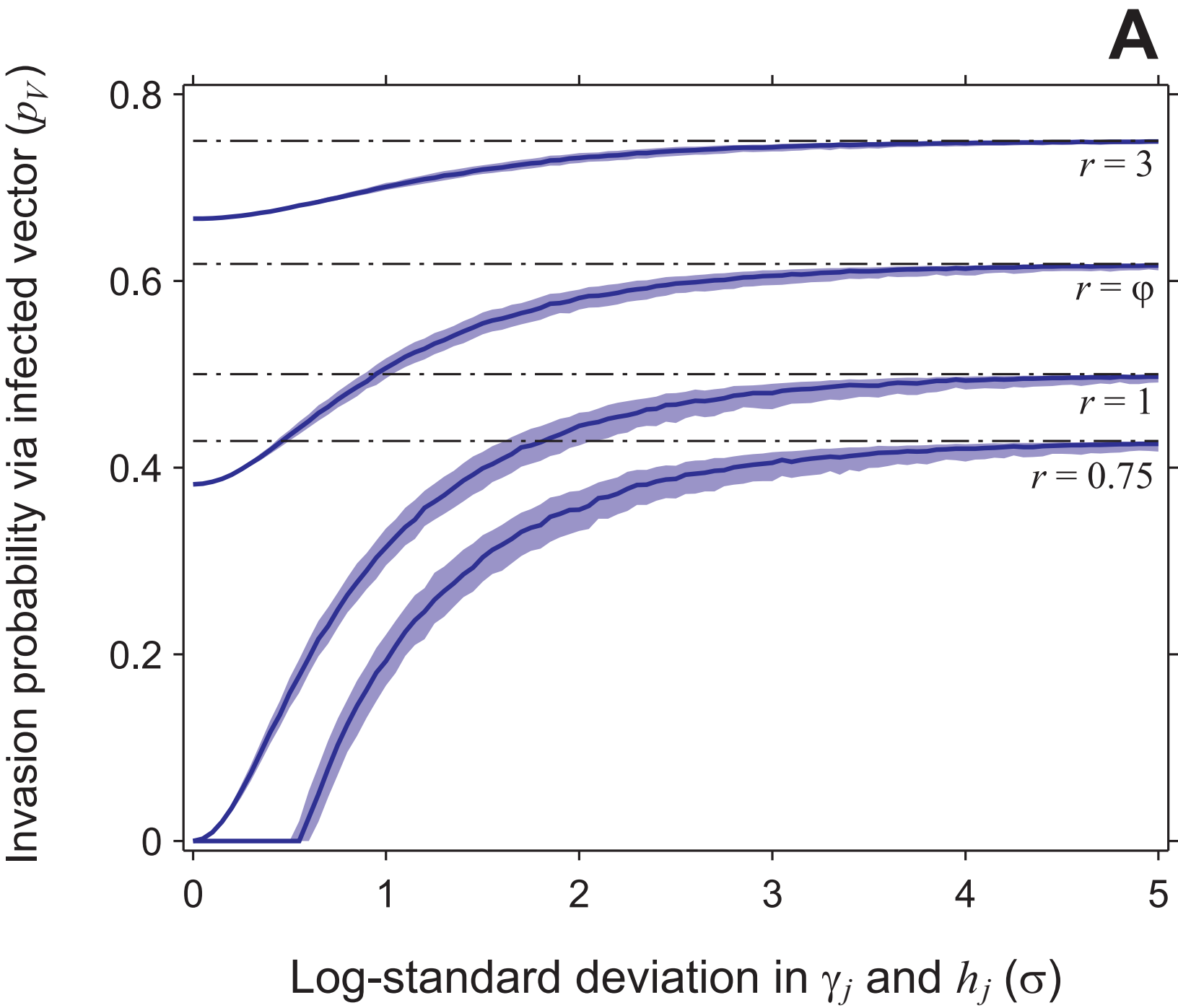
4. Figure 2

Click here to download 4. Figure: fig2.eps



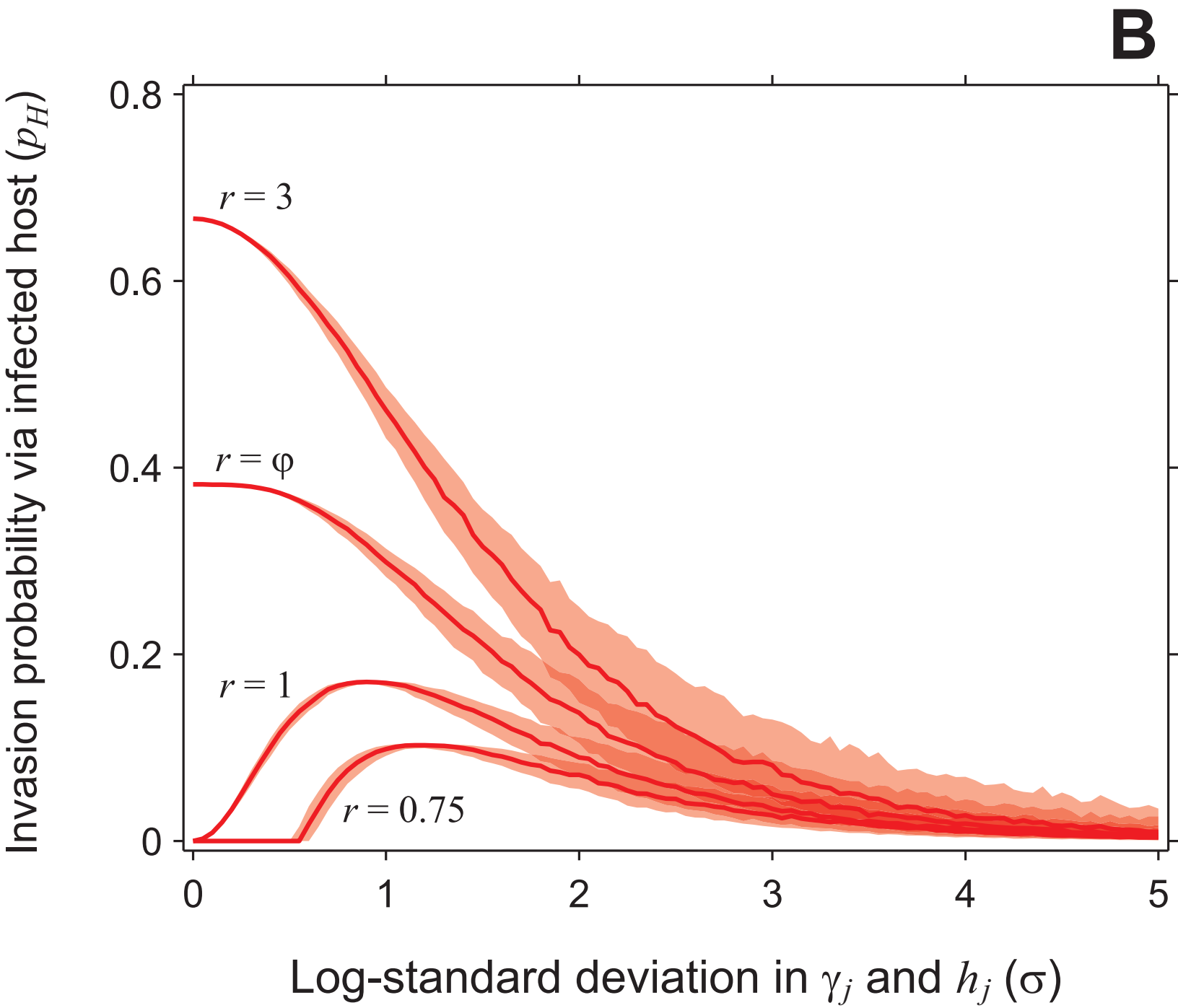
4. Figure 3a

Click here to download 4. Figure: fig3a.eps



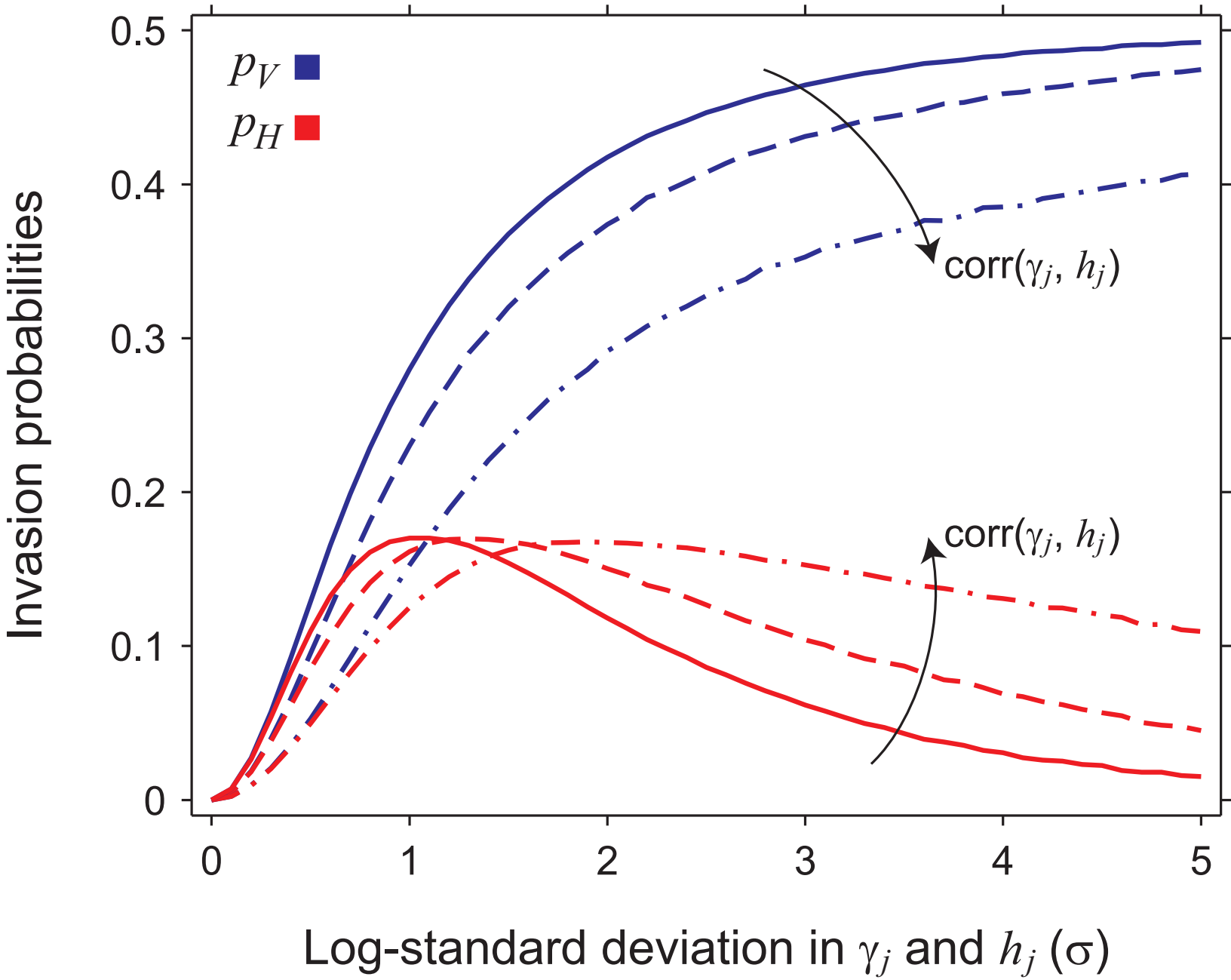
4. Figure 3b

Click here to download 4. Figure: fig3b.eps



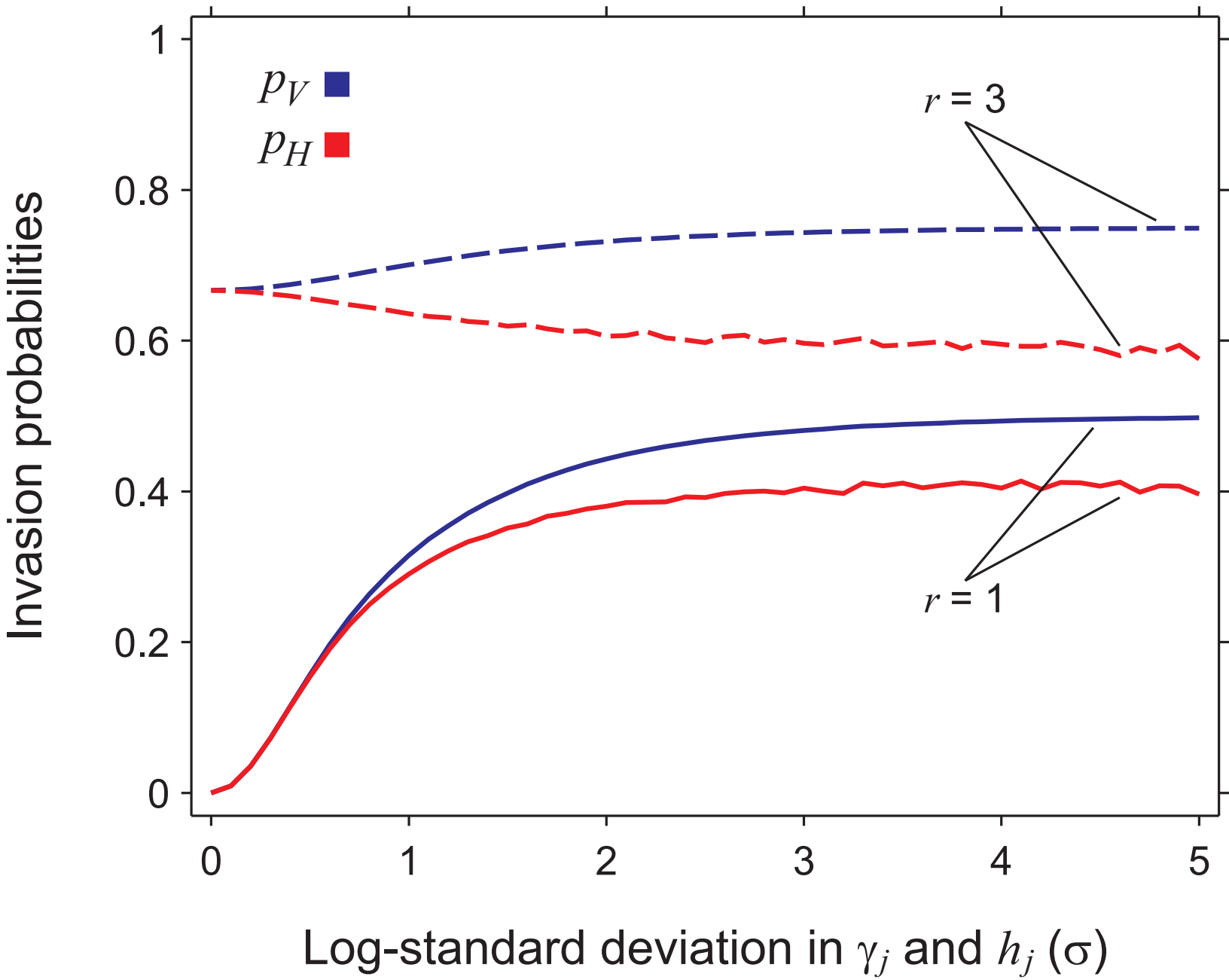
4. Figure 4

Click here to download 4. Figure: fig4.eps



4. Figure 5

Click here to download 4. Figure: fig5.eps



6. Supplementary Material for on-line publication only

Click here to download 6. Supplementary Material for on-line publication only: SuppMat_revision_JTB.doc