The Effect of Heterogeneity on Invasion in Spatial Epidemics: From Theory to Experimental Evidence in a Model System

Franco M. Neri, Anne Bates, Winnie S. Fuchtbauer, Francisco J. Perez-Reche, Sergei N. Taraskin, Wilfred Otten, Douglas John D. J. Bailey, Christopher A. Gilligan

HAL Id: hal-02646741
https://hal.inrae.fr/hal-02646741

Submitted on 29 May 2020

The Effect of Heterogeneity on Invasion in Spatial Epidemics: From Theory to Experimental Evidence in a Model System

Franco M. Neri[1]*, Anne Bates[1], Winnie S. Füchtbauer[1,2], Francisco J. Pérez-Reche[3], Sergei N. Taraskin[4], Wilfred Otten[5], Douglas J. Bailey[6], Christopher A. Gilligan[1]

1 Department of Plant Sciences, University of Cambridge, Cambridge, United Kingdom, 2 Department of Molecular Biology, Aarhus University, Aarhus, Denmark, 3 Department of Chemistry, University of Cambridge, Cambridge, United Kingdom, 4 St. Catharine's College and Department of Chemistry, University of Cambridge, Cambridge, United Kingdom, 5 The SIMBIOS Centre, University of Abertay Dundee, Dundee, United Kingdom, 6 INRA UMR1099 BiO3P (Biology of Organisms and Populations Applied to Plant Protection), Le Rheu, France

Abstract

Heterogeneity in host populations is an important factor affecting the ability of a pathogen to invade, yet the quantitative investigation of its effects on epidemic spread is still an open problem. In this paper, we test recent theoretical results, which extend the established "percolation paradigm" to the spread of a pathogen in discrete heterogeneous host populations. In particular, we test the hypothesis that the probability of epidemic invasion decreases when host heterogeneity is increased. We use replicated experimental microcosms, in which the ubiquitous pathogenic fungus *Rhizoctonia solani* grows through a population of discrete nutrient sites on a lattice, with nutrient sites representing hosts. The degree of host heterogeneity within different populations is adjusted by changing the proportion and the nutrient concentration of nutrient sites. The experimental data are analysed via Bayesian inference methods, estimating pathogen transmission parameters for each individual population. We find a significant, negative correlation between heterogeneity and the probability of pathogen invasion, thereby validating the theory. The value of the correlation is also in remarkably good agreement with the theoretical predictions. We briefly discuss how our results can be exploited in the design and implementation of disease control strategies.

Introduction

Host heterogeneity is receiving increasing attention as one of the factors affecting the dynamics of epidemic spread. The properties of individual hosts, such as contact rate, susceptibility, or infectiousness, can vary across a population as a result of environmental [1,2], genetic [3] and immunogenic [4] factors. Such variability is typically difficult to measure empirically, and has been successfully quantified only in a few significant cases, concerning plant [5], animal [6–10], and human diseases [6,7,11]. Even more important, a few studies [6,8–10,12] succeeded in addressing a key epidemiological question: what is, if any, the effect of individual variability on the risk of *epidemic invasion* [13] (that is, the chance that a pathogen, starting from a single or a few infected hosts, will be able to infect a significant proportion of the whole population). For example, it was found that variations in prevalence of *E. coli* O157 among cattle populations were best explained by individual variability in bacterial load and infectiousness[8]; in plant populations, the rate and pattern of disease invasion were found to be influenced by variations in individual susceptibility and transmission rates [2,14]. An important consequence of these findings is that variability can affect *invasion*

thresholds, i.e., the critical values of the parameters of the system (transmission rate, host density, etc.) that determine whether or not a pathogen can invade [13,15]. Since invasion thresholds are a central idea underlying most control strategies [13], the practical implications are huge: it is known that control strategies can benefit from the knowledge of host variability [11,12]; on the other hand, such strategies can fail if variability is ignored and "averaged out" [6,7,12], due to serious misestimation of the parameters of the epidemic model [14]. However, despite such implications, and a continuing effort to explore the problem with theoretical models (e.g., for fully mixed populations [16–20], metapopulations models [21,22], and complex networks [23–25]), rigorous experimental testing has been limited, restricting our understanding of the problem.

The experimental results presented here test for the first time the existence of a link between host heterogeneity and epidemic thresholds in a broad, relevant class of spatially-extended systems, thereby confirming recent theoretical predictions [26]. The class comprises those systems where the pathogen is transmitted between neighbouring hosts: such mode of transmission is typical of many diseases, such as soil-borne diseases in plant populations [27–29]; plant pathogens spreading among neighbouring fields or

Author Summary

Pathogen spread and epidemic invasion in plant, animal and human populations depend on host properties (infectivity, susceptibility) that can vary amongst hosts within the same population. However, such host variability (or heterogeneity) is typically difficult to control experimentally, and little explicit research has been done on its effects on pathogen invasion. We present the first systematic investigation on the spread of a pathogen (the fungal plant pathogen *Rhizoctonia solani*) in repeatable microcosm populations (nutrient sites, representing hosts, on a lattice) with varying heterogeneity. Recent theoretical models for heterogeneity make the following prediction, that we set as our hypothesis: increasing heterogeneity (keeping all other properties fixed) results in a decrease of the probability of pathogen invasion. We validate the hypothesis using a combination of statistical methods. Moreover, we find that the decrease in probability of invasion is also in good quantitative agreement with the theoretical predictions. We discuss how our results can be exploited for disease control strategies that increase the heterogeneity of a system in order to limit pathogen spread.

farms [30]; animal pathogens spreading within populations of hosts living in a fixed habitat [31,32]. A "percolation framework" [33], commonly used to describe these systems, is adopted here to model epidemic spread and invasion. Percolation theory provides conceptual tools that allow "scaling up" from pathogen transmission at the small (between-host) scale to epidemic invasion at the large (population) scale [27]. The use of this framework for the identification of invasion thresholds has been experimentally validated in a few remarkable cases [27,28,31,32]. Many diseases characterized by short-range transmission are also well described by an SIR (susceptible–infected–removed) stochastic model, where an infected host (I) can transmit the pathogen to its susceptible (S) neighbours for some interval of time, after which it is permanently recovered or removed (R); the probability that transmission actually occurs before removal is the *transmissibility*.

Previous experiments [27,28], using replicable microcosm lattice systems as models for SIR soil-borne plant diseases, succeeded in validating two key predictions from percolation theory [34]: that invasion and threshold behaviour for SIR diseases are controlled by the transmissibility; and that the epidemic threshold value of the transmissibility coincides within experimental precision with the "bond-percolation" threshold for the system ([33]; see Text S1 for details). Hence, a pathogen spreading on a lattice will never invade the population when the transmissibility is lower than the bond percolation threshold, while there will be a risk of invasion when the transmissibility is higher [27,34].

The experiment described in the present paper is inspired by a model by Neri *et al.* [26] that goes beyond the percolation-based theory of SIR epidemics by including host variability in the probability of transmission. In a homogeneous system, all the hosts (once infected) are able to transmit the pathogen with the same probability. Conversely, a heterogeneous system, typical of most natural host populations, can be modelled by assuming that the transmissibility is not constant across the population, but is a random variable, drawn for each host from a given common distribution [26]. Epidemic invasion in such systems can then be characterized by two parameters: the average value of the

transmissibility over the population and its variance. The variance is used as a measure of the heterogeneity of the system. It was found [26] that both the average and variance of the transmissibility contribute to invasion, but with opposite effects: while the former increases the probability of invasion, the latter leads to a decrease (Figure 1A). As a consequence, invasion can be described by a phase diagram in the two-dimensional parameter space for the average and variance of the transmissibility (Figure 1B). The phase diagram contains two distinct regions: a region with low average transmissibility and large variance, where an epidemic will never invade the system (non-invasive regime), and a region with large average transmissibility and low variance, where the epidemic can invade (invasive regime). Instead of a single threshold value for the transmissibility, a threshold curve (phase boundary) separates the two regimes. A further important result is that, in a given region of the phase diagram (Figure 1B), it appears to be possible to "switch" a system from being invasive to being non-invasive (or *vice-versa*) by keeping the average transmissibility constant and changing only the variance, i.e., by changing only the heterogeneity of the system. The latter result is the main motivation for the present experiment.

We use replicable microcosms [27,28] to test the predictions of the heterogeneous SIR model[26]. We take advantage of an experimentally validated paradigm [27,28], whereby the spread of an SIR epidemic in a discrete host population is equivalent to, and can be investigated by, the spread of fungi by mycelial growth among a population of nutrient sites. We analyse the growth of the ubiquitous pathogenic fungus *Rhizoctonia solani* in microcosm populations composed of nutrient sites arranged on a lattice. Here, the term "transmissibility" (which in this particular case is related to infectiousness of donor sites, see Text S1) represents the probability of fungal spread from one site to another. Our main aim is to answer the following questions: does host heterogeneity (measured by the variance of the transmissibility) affect the probability of invasion, and if this is the case, how? We also ask: is it possible to see an effect of heterogeneity on the threshold for invasion of the system? Since our experiments are conducted on relatively small populations, while tresholds are rigorously defined only for infinite systems (Figure 1), we also address the question: can thresholds for invasion be estimated from a small-scale experiment?

We set up a series of notional experimental treatments (replicated populations), designed in such a way to ensure an appropriate range for the average and variance of the transmissibility. The notional values of the parameters are chosen according to the theoretical predictions of Neri *et al.* [26]: the average transmissibility is the same for all the populations, while the variance differs amongst treatments (cf. the dash-dotted line in Figure 1D). This design allows us to determine how the probability of invasion, calculated using spatio-temporal experimental maps, changes with the variance. In practice, because of inherent variability of the systems, replicates within the same treatment differ. Accordingly, at the end of the experiment, we re-estimate the values of the average and variance of the transmissibility for *each* individual replicate from spatio-temporal maps, using Bayesian Markov-chain Monte Carlo (MCMC) methods [35]. The new estimated parameters are then used for a statistical analysis carried out on the pooled set of all the populations. The pooled analysis proves to be effective in assessing the joint contribution of average and variance of the transmissibility to the probability of invasion.
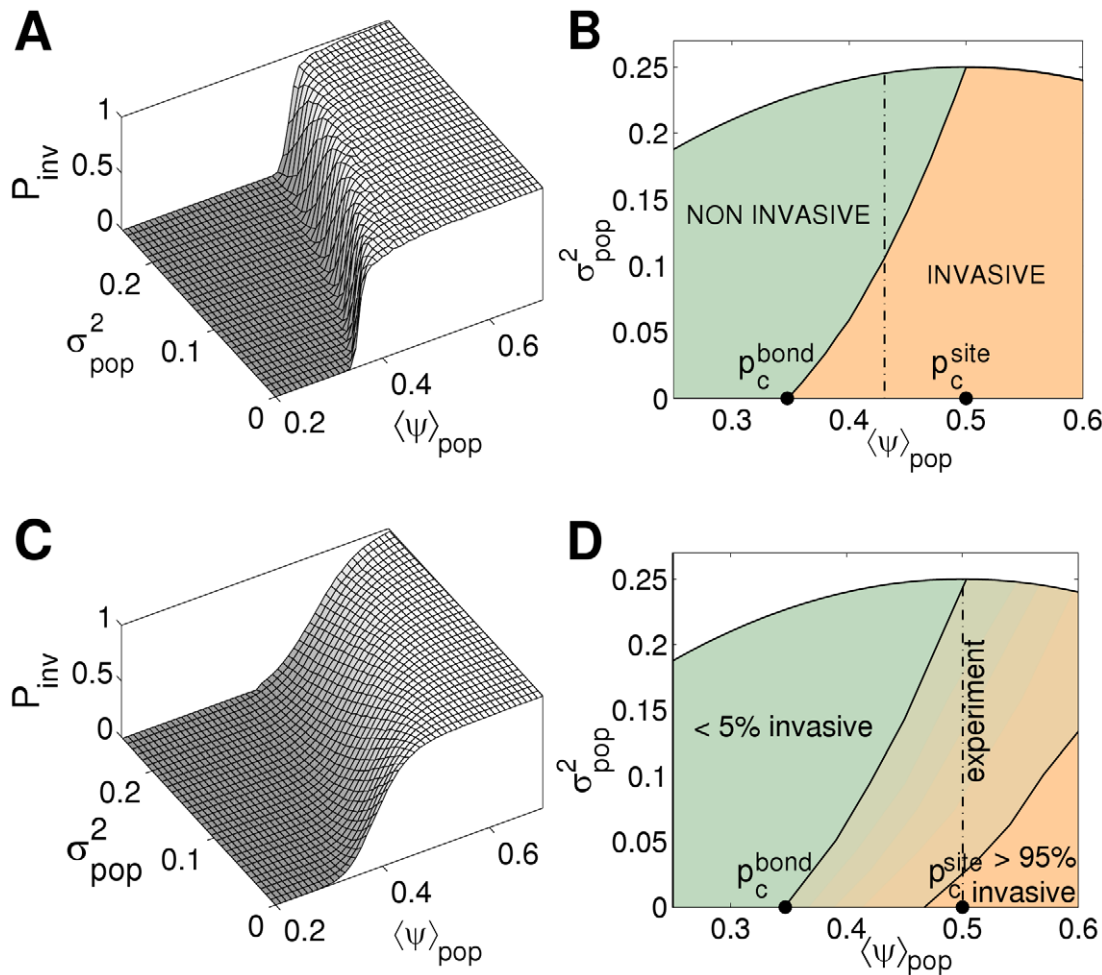
Figure 1. Epidemic invasion in heterogeneous populations: invasion probability and phase diagram. We consider systems of two different sizes, on triangular lattices with the same topology as for the populations used in the experiment. The graphs were obtained with numerical simulations (see Text S1 for details). The probability of epidemic invasion, P_{inv}, is studied as a function of the mean, $\langle\psi\rangle_{pop}$, and the variance, σ^2_{pop}, of the transmissibility over the system. (A) Probability of invasion for the larger system (24031 sites) as a function of the mean and variance of the transmissibility. (B) Phase diagram for invasion, calculated numerically from (A). The solid line marks the phase boundary where the transition between the invasive and the non-invasive regimes occurs (corresponding to the sharp jump in (A)). The quantities $p_c^{bond} \simeq 0.347$ and $p_c^{site} = 0.5$ are the bond- and site-percolation thresholds, respectively, for the triangular lattice. The white region beyond the parabolic thick curve corresponds to combinations of values of $\langle\psi\rangle_{pop}$ and σ^2_{pop} that are theoretically impossible. The large-size phase boundary is shown here to exemplify the (more rigorously defined) phase boundary for infinite-size systems (see discussion in Text S1), for which it provides a good numerical approximation. (C) P_{inv} for the smaller system (217 sites). (D) Effective phase diagram for invasion in the small-size system calculated from (C): as expected for such systems, the transition between the two regimes is "smeared out" across a broader region (delimited by solid lines corresponding to the values $P_{inv} = 0.05$ and $P_{inv} = 0.95$). For systems of both sizes, it is possible to change invasion regime by changing the variance σ^2_{pop} only, with the average $\langle\psi\rangle_{pop}$ kept constant (dash-dotted lines in (B) and (D); the position of the line for the two systems is different owing to finite-size effects).
doi:10.1371/journal.pcbi.1002174.g001

Results

Invasive spread and individual rates in heterogeneous populations

Six experimental treatments, labelled from A to F, were designed (see below and Table S1), with 30 replicated populations for each treatment. Each treatment corresponded to a population of 217 nutrient sites (agar dots) arranged on a triangular lattice, comprising a fraction, ρ, of "occupied" sites (randomly selected to be occupied with nutrient), the remaining fraction $1-\rho$ being left empty. Henceforth, the symbol ψ will used to denote transmissibility in general; ψ_{site} for the transmissibility of a site with nutrient; $\langle\psi\rangle_{pop}$ and σ^2_{pop} for the mean and variance, respectively, of the transmissibility over a population. The value of ψ_{site} depends on the amount of nutrient (see Text S2 for the

determination of ψ_{site}). The population mean and variance of the transmissibility of the experimental populations are given by:

$$\langle\psi\rangle_{pop} = \rho\,\psi_{site} \tag{1a}$$

$$\sigma^2_{pop} = \rho(1-\rho)\,\psi^2_{site}, \tag{1b}$$

and were controlled by adjusting ρ and ψ_{site} in order to keep $\langle\psi\rangle_{pop} \simeq 0.5$ approximately constant for all treatments (cf. the dash-dotted line in Figure 1D), while σ^2_{pop} was increased by regular intervals, in alphabetical order, from A (homogeneous system) to F (maximally heterogeneous system; see Text S1 for more details).

Spatio-temporal maps of fungal colonisation dynamics were used, in order to count the cumulative number of colonised sites over time, and to identify those replicates in which the fungus spreads invasively (Figure 2). For each treatment, a fraction of the replicates had to be discarded because of contamination from external sources, leaving a total of 151 replicated populations out of the initial 180. The fraction of invading replicates (Figure 2) shows that the probability of invasion (P_{inv}) decreased from treatments A to F, following the predicted trend (Table S1). Going from A to F, the fractions of available sites (i.e., occupied by nutrient) reached by the colony also decreased (with the exception of treatments E and F, probably due to stochasticity).

The variability in the final number of colonised sites amongst replicates of the same treatment (Figure 2) can be attributed to two distinct factors: (1) it was partly an effect of the stochastic nature of the colonisation process, which is taken into account by our model, but (2) it could also be caused by variations in the value of ψ_{site} amongst different replicates, due to uncontrollable factors such as variations in the environmental conditions amongst replicate populations contained within different Petri plates. Within-treatment variation of ψ_{site}, which is not accounted for in our model, can also significantly change the probability of fungal invasion. For this reason, re-assessment of ψ_{site} was conducted for each individual replicate in order to test within-treatment variation. In what follows, we call $\psi_{site}(T,r)$ the transmissibility for replicate r of treatment T ($T=A,\ldots,F$), and $\widehat{\psi}_{site}(T,r)$ the corresponding estimate (see Materials and Methods for definitions and an explicit example).

In order to estimate ψ_{site}, we modelled the time evolution of the probability of transmission between nutrient sites with a Weibull function (see Materials and Methods for details), initially characterized by a single rate of spread. However, preliminary inspection of the population data showed *two* distinct stages for the fungal colony spread: an initially slower process (first stage), followed by a transition to a faster process (second stage). This behaviour can be explained as an effect of nutrient translocation, common to several fungal species [36,37]: mycelial colonies growing from different nutrient sites are able to share resources, so that their rate of spread increases with the number and connectivity of colonised sites in the system. We accounted for this effect by using a modified Weibull model: the new model included two distinct rates, corresponding to the two stages of the process, and the "switching" time of the transition from the slower to the faster stage. We found that the two-rate model could parametrise the data very efficiently, and provided a good estimation of the posterior distribution for ψ_{site} (see Figures S2–4 in Text S3).

The analysis of posterior distributions for $\psi_{site}(T,r)$ for different replicates, r, of the same treatment T, revealed considerable differences from the notional values used for the experimental design (summary statistics in Table S1; see Text S3 for the complete set of results). In particular, $\widehat{\psi}_{site}(T,r)$ showed a systematic shift to lower values of ψ for treatments D, E, F. On average, $\widehat{\psi}_{site}(T,r)$ *between* treatments still increased from treatment A to F. However, and most important, within-treatment heterogeneity of $\psi_{site}(T,r)$ was in general large (Table S1). Such
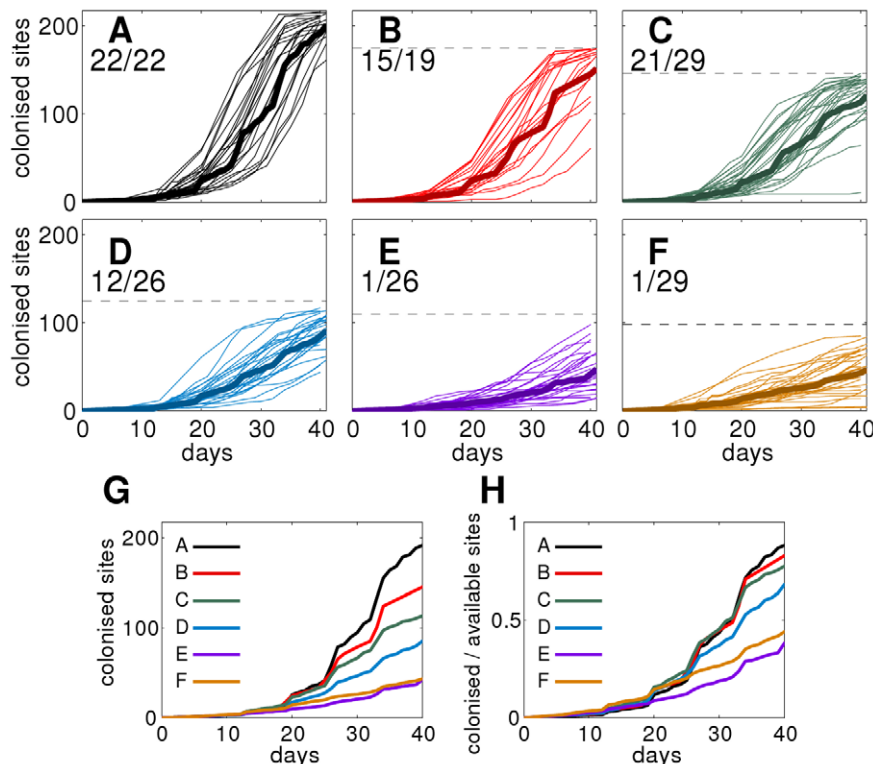
Figure 2. Experimental colonisation curves for the six different treatments. (A–F) Individual colonisation curves for each of the replicates of a treatment (thin lines) and the average over all the replicates (bold solid line). The upper limit of the vertical axis in all the panels concides with the total number of sites in the population (i.e., 217), while the horizontal dashed line in treatments B to F marks the number of available sites (i.e., occupied by nutrient). The fraction of invading replicates per treatment is displayed in each panel as "number of invading replicates divided by total number of replicates". (G) Comparison of average colonisation curves for all the treatments. (H) Comparison of average colonisation curves, normalised to the total number of occupied sites per treatment.
doi:10.1371/journal.pcbi.1002174.g002

heterogeneity can be modelled explicitly using a hierarchical model [35] (Text S3). For the purpose of the present experiment, however, the main outcome of this analysis is that deviations from the initial, notional values of ψ_{site} (hence, of $\langle\psi\rangle_{\text{pop}}$ and σ^2_{pop}) within each treatment are significant. Hence, instead of the nominal values of the parameters, we used the re-estimated values for the final statistical analysis (Figure 3), which was carried out on the pooled set of all the populations. For each population, $\widehat{\psi}_{\text{site}}(T,r)$ was used to obtain new estimates for the mean and variance of the transmissibility, $\langle\widehat{\psi}\rangle_{\text{pop}}(T,r)$ and $\widehat{\sigma}^2_{\text{pop}}(T,r)$ (represented by circles and crosses, respectively, in the $\left(\langle\psi\rangle_{\text{pop}}, \sigma^2_{\text{pop}}\right)$ plane shown in Figure 3).

Experimental phase diagram for invasion

The data presented in Figure 3 provide evidence for the main hypothesis of the paper: namely, that the variance as well as the mean influence the probability of invasion. Two different statistical tests were used to test the hypothesis. Linear discriminant analysis (LDA) was used to find a linear separatrix between the invasive and non-invasive regimes in the parameter space (solid line in Figure 1C; discriminant function $-3.56 + 10.46\langle\psi\rangle_{\text{pop}} - 9.62\sigma^2_{\text{pop}}$). The function classified correctly 75 out of the 79 non-invasive points in the graph (95% success rate) and 49 out of the 72 invasive points (68% success rate). Goodness-of-fit tests gave Wilks' $\Lambda = 0.503$, $p < 0.001$. Remarkably, the linear separatrix approximates very well the theoretical prediction for the phase boundary in an infinite system (see Text S1), shown by the dash-dotted line in Figure 3 (and approximated by the solid line for the large system in Figure 1B).

A multiple logistic regression test (function $Y = \beta_0 + \beta_1\langle\psi\rangle_{\text{pop}} + \beta_2\sigma^2_{\text{pop}}$, with $Y = \ln\left(P_{\text{inv}}/(1 - P_{\text{inv}})\right)$) was also conducted, motivated by the fact that the theoretical 3D invasion curves (Figure 1) can be well fitted by a multivariate logistic model. The values of $\langle\psi\rangle_{\text{pop}}(T,r)$ and $\sigma^2_{\text{pop}}(T,r)$ for each population were weighted with the reciprocal of the variance, calculated from the corresponding posterior curves. The test yielded the estimated parameter values (95% confidence intervals are indicated): $\beta_0 = -9.4 \pm 1.9$, $\beta_1 = 29.6 \pm 5.4$, $\beta_2 = -25.8 \pm 5.9$ ($p < 0.001$ for all the parameters). The coefficient ratio $-\beta_1/\beta_2 = 1.15 \pm 0.47$ is compatible with the slope 1.09 found with LDA. Both tests show the statistical significance of σ^2_{pop} as a predictor for P_{inv}.

Discussion

We have shown experimentally that between-host variability affects the nearest-neighbour spread of a pathogen in a population: when the variability is increased, the probability of epidemic invasion decreases. From a broad point of view, our results answer a very general question: what is the effect of individual variability on disease spread? [6,7,38], the answer being valid for systems with a locally spreading pathogen. We have exploited a percolation-based approach, which is widely used for such systems, and which has been experimentally tested for disease invasion in plant [27,28] and animal populations [31], under the assumption that these populations were homogeneous. Theoretical work [26,39] has shown how to include heterogeneity in the percolation paradigm, by introducing a quantitative measure of between-host variability [26] (defined as the variance, σ^2_{pop}, of the transmissibility ψ within a population). This way, it becomes possible to make quantitative predictions on pathogen invasion in the presence of heterogeneity.

Here, for the first time, the approach and quantitative predictions of Neri *et al.* [26] have been validated experimentally. We have exploited the saprotrophic spread of *R. solani*, a ubiquitous plant pathogen, in simplified microcosms of hosts represented by agar dots arranged on regular lattices. Previous work showed that such simple microcosm systems are representative of epidemic systems
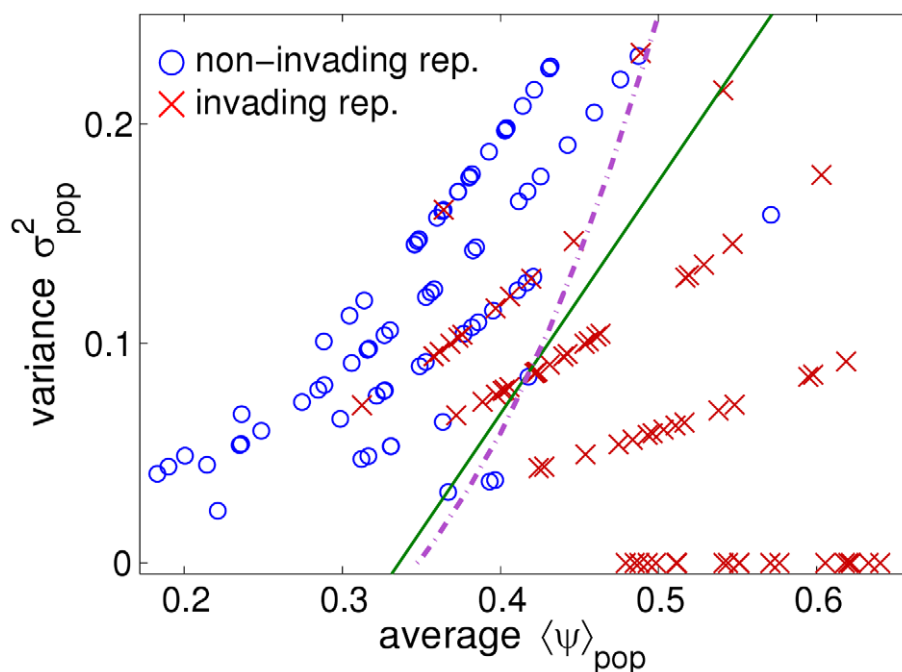
Figure 3. Estimated population parameters and linear discriminant analysis for invasion in the plane $\left(\langle\psi\rangle_{\text{pop}}, \sigma^2_{\text{pop}}\right)$**.** Red crosses and blue circles correspond to invasive and non-invasive replicates, respectively (error bars not shown here). The green thick line is the discriminant function separating the invasive and non-invasive regimes. The purple dash-dotted line is the phase boundary for an infinite system with the same topology (see Text S1), and is close to the solid line for the large-system phase boundary in Figure 1B.
doi:10.1371/journal.pcbi.1002174.g003

that can be described by SIR spatial models [27,28]. We have been able to show quantitatively (Figure 3) the effect of within-population heterogeneity (σ^2_{pop}) on the probability of epidemic invasion, P_{inv}, and demonstrated the existence of a statistically significant, negative correlation between P_{inv} and σ^2_{pop}. At the same time, our results and analysis show that the phase diagram for invasion (in principle, defined only for infinite systems), which includes the effects of the variance, can be well approximated using small-scale experiments.

Our analysis showed that within-treatment variability can be large enough to mask the effects of experimental treatments in replicated populations (see Table S1 and Text S3). The methodology we adopted is relevant, in general, to the case when experimental factors are subject to environmental stochasticity, thus are not under the full control of the experimenter. Specifically, while one of our experimental factors (the fraction ρ of occupied sites in a population) was known exactly, the other (ψ_{site}) could vary considerably amongst different replicates of the same putative treatment (Figures S2 and S3 in Text S3). Such within-treatment variability is not always necessarily relevant (e.g., in [27,28]), but it can interfere with the statistical analysis when the values of the experimental parameters need to be known with high precision, as in our case. We showed that within-treatment variability can be efficiently assessed, via MCMC Bayesian techniques, with a *post-hoc* estimation of ψ_{site} from each individual replicate. The parameter estimation step has also provided an efficient tool to overcome the difficulties due to within-treatment variability. It was indeed possible to analyse the pooled set of the replicates from all the treatments (Figure 3) instead of the average response of a treatment: this approach has proved to be successful in estimating the phase boundary for invasion.

Variability and implementation of control strategies

Our results have a potentially high impact in finding control strategies for the spread of disease. Let us consider a homogeneous system, with the same topology as for our experimental microcosms (Figure 4), but where all the sites have initially a high transmissibility ($\psi_{\text{site}}=1$, yielding a probability of pathogen invasion $P_{\text{inv}}=1$). Assume that an epidemic is about to start from the central site, and we can control disease spread by applying, only once, a control agent (a protectant) to all or part of the sites. The effect of the agent on a site is (for simplicity) linear, so that an amount c of control agent brings ψ_{site} to the value $\psi_{\text{site}}(1-c/c_{\text{max}})$, where c_{max} is the amount needed to make the site non-infectious (i.e., $\psi_{\text{site}}=0$). If the amount of control agent at our disposal is fixed and less than $N c_{\text{max}}$ (which would be needed to make all the sites non-infectious), the question is how best to allocate such amount amongst the sites (a similar problem is discussed in [26]).

In particular, if we assume that the amount of control agent is equal to $N c_{\text{max}}/2$, different allocation strategies correspond to our experimental treatments. We could opt for a homogeneous strategy, applying an amount $c_{\text{max}}/2$ of agent to each individual site, and reducing its transmissibility from $\psi_{\text{site}}=1$ to $\psi_{\text{site}}=0.5$ (exemplified by treatment A, Figure 4A). Alternatively, we could apply a heterogeneous strategy: for example, selecting a fraction $\rho=3/4$ of the sites to be treated with an amount $c_{\text{max}}/3$ (bringing their transmissibility to $\psi_{\text{site}}=0.66$), and the remaining fraction (1/4) of the sites to be treated with c_{max} and made non-infectious (exemplified by treatment C, Figure 4C). For any strategy, the value of the final mean transmissibility is always $\langle\psi\rangle_{\text{pop}}=0.5$, and only σ^2_{pop} changes. The results presented in this paper show that, if the cost of any strategy is constant (i.e., it depends only on the amount of control agent applied), the maximally heterogeneous
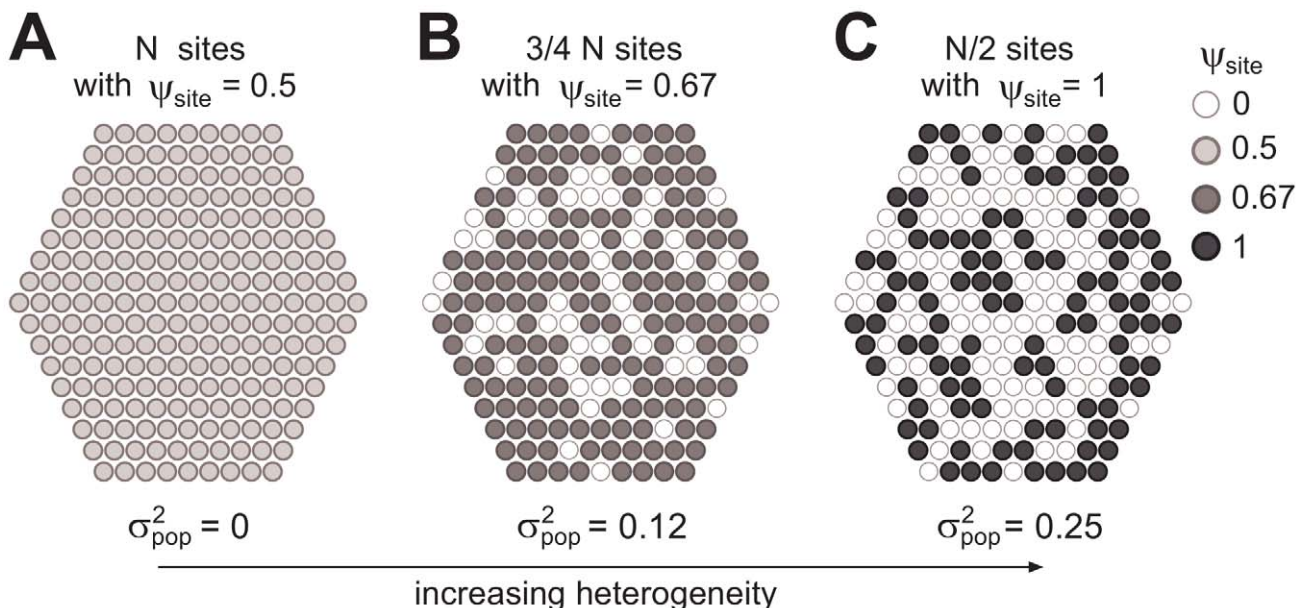

Figure 4. Three experimental treatments for heterogeneous populations of nutrient sites. The treatments presented here are slightly idealized versions of the real experimental treatments A, C, F (the condition $\langle\psi\rangle_{\text{pop}}=0.5$ is used here instead of $\langle\psi\rangle_{\text{pop}}\simeq0.5$). The fraction ρ of sites with nutrient and the transmissibility ψ_{site} are ajdusted so that $\langle\psi\rangle_{\text{pop}}=0.5$ (given by Equation 1) for all the treatments, while heterogeneity (measured by σ^2_{pop} and given by Equation 2) increases. (A) Treatment A: homogeneous system, with all the N sites occupied ($\rho=1$) and $\sigma^2_{\text{pop}}=0$. (B) Treatment C, heterogeneous system with $\rho\simeq0.67$ and $\sigma^2_{\text{pop}}=0.12$. (C) Treatment F: heterogeneous system with $\rho=0.5$ and $\sigma^2_{\text{pop}}=0.25$ (the "maximally heterogeneous" treatment that maximising the variance given the mean $\langle\psi\rangle_{\text{pop}}=0.5$).
doi:10.1371/journal.pcbi.1002174.g004

strategy (exemplified by treatment F, Figure 4C) gives the maximal decrease of the probability of invasion in every finite system (see [26] for a discussion of analogous results for infinite systems).

Materials and Methods

Experimental design

In the experimental design and the subsequent analysis, the parameters were evaluated in two steps, before and after the population experiment. Before the population experiment, the dependence of ψ_{site} on the agar dot nutrient concentration was estimated by means of "placement" experiments involving only pairs of sites [27,28] (see Text S2 for the details of the placement experiments and the estimation of ψ_{site}). The results of pair experiments were used to select the values of the parameters ρ and ψ_{site} for the notional treatments of the population experiment (Table S1; three of the treatments are exemplified in Figure 4). The real values of ψ_{site}, however, can be affected by environmental conditions, and can change significantly amongst replicates of the same treatment. Therefore, after the end of the experiment, new estimates of ψ_{site} were obtained for each population individually, using Markov-chain Monte Carlo (MCMC) methods, in order to assess within-treatment variability. The new estimates were then used for the pooled analysis shown in Figure 3.

In our experimental systems, each agar dot consisted of a small aliquot (10 μL, 3 mm diameter) of potato dextrose agar (PDA), with nutrient concentrations ranging from 3% to 10%. Sites of agar dots were spotted onto a triangular lattice in large Petri plates (140-mm diameter) at 8-mm apart (from centre to centre). Each treatment was replicated 30 times using independent randomisation schemes, leading to a total of 180 populations. The central agar site of each population was inoculated with a single hyphal strand removed from the growing edge of a 4d-old colony of *R. solani* R5 (AG 2-1) grown on water agar. Moist filter paper was placed in the lid of each Petri plate to avoid desiccation of the agar and the plates were sealed and incubated in the dark at 23°C and assessed for 41d using a binocular microscope (40x), recording the number and locations of colonised sites. For each treatment, 20 replicates were assessed every 2 days, and the remaining 10 replicates were assessed weekly. Spatio-temporal maps (snapshots of colonisation over time) were therefore produced.

Criteria for invasion. The fact that the transition between the invasive and the non-invasive regime is "smeared out" (Figure 1C) raises the question of how accurately the transition can be approximated in small experimental microcosms. This depends in a crucial way on the criterion used to define invasion. Previous authors [27,28] considered reaching *at least one edge* of the system boundaries at the end of the experiment (starting from a single infected site at the center) as equivalent to invasion, on the basis that the epidemic is certainly non-invasive when no edge is reached. In this paper, we adopt a different criterion: invasion occurs when *all the edges* are reached by the pathogen at the end of the experiment (i.e., six edges in our case, which is the criterion used in Figure 1). Our choice was made after comparing the probability of invasion for simulated epidemics calculated with the two different criteria (results not shown here). We found that (i) for large system sizes (cf. Figure 1A) the values of P_{inv} calculated with the two criteria tend to coincide, but (ii) the deviation in probability between large and small systems (Figure 1A versus Figure 1C) is systematically smaller when the "six-edges" criterion is used. Hence, the latter criterion gives a better prediction of the large-scale behaviour of an epidemic from its small-scale behaviour.

Fungal spread, parameter estimation and data analysis

The value of ψ_{site} was re-estimated for each individual population at the end of the experiment, using an MCMC method [35]. The growth of the fungal colony between two neighbouring sites was modelled as a time-inhomogeneous Poisson process [40] described by a Weibull distribution multiplied by the transmissibility:

$$f_W(t) = \psi_{\text{site}}\left(t^{k-1}/\lambda^k\right)e^{-(t/\lambda)^k} \tag{2a}$$

$$F_W(t) = \psi_{\text{site}}(1 - e^{-(t/\lambda)^k}), \tag{2b}$$

where $f_W(t)$ is the distribution of colonisation times, and $F_W(t)$ is the probability of colonisation as a function of t; λ is the time scale of the process and k is a shape parameter. In order to account for the observed transition in rates (slower spread at the beginning, faster spread towards the end of the experiment), we introduced a "switching time" t_{sw}, such that $\lambda = \lambda_1$ for $t \leq t_{\text{sw}}$ and $\lambda = \lambda_2$ for $t > t_{\text{sw}}$ (in general, $\lambda_2 < \lambda_1$). Thus, the model has 5 parameters,
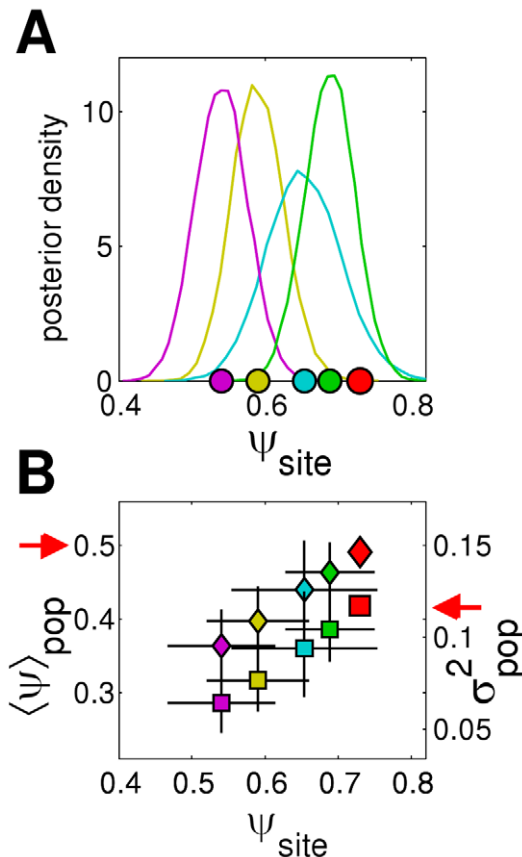

Figure 5. Within-treatment variation of ψ_{site}, re-estimated with Monte Carlo methods. (A) Posterior distributions for ψ_{site} for four different replicates of treatment C (coloured lines). The means of the posteriors (coloured circles) shift away from the value $\psi_{\text{site}} = 0.73$ (red circle) used for the experimental design (Table S1). (B) The mean of the posterior is used to re-calculate $\langle\psi\rangle_{\text{pop}}$ and σ^2_{pop} for each replicate (coloured diamonds and squares, respectively), using the 95% credible interval as the error bar. The new values differ from the nominal values $\langle\psi\rangle_{\text{pop}} = 0.49$ and $\sigma^2_{\text{pop}} = 0.12$ (red diamond and square, respectively, marked by arrows) used in the experimental design (Table S1).
doi:10.1371/journal.pcbi.1002174.g005

represented by the vector $\boldsymbol{\theta} = (\psi_{\text{site}}, \lambda_1, \lambda_2, k, t_{\text{sw}})$. In order to estimate the parameters, we adopted a Bayesian framework, treating the parameters as random variables themselves. The posterior distribution for $\boldsymbol{\theta}$ given the observed data D, $\pi(\boldsymbol{\theta}|D)$, is given by Bayes' formula $\pi(\boldsymbol{\theta}|D) \propto \Pr(D|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$, where $\pi(\boldsymbol{\theta})$ is the prior distribution of the parameters (reflecting our initial belief in their values), and $\Pr(D|\boldsymbol{\theta})$ is the likelihood (the probability of the observed data given $\boldsymbol{\theta}$). The posterior $\pi(\boldsymbol{\theta}|D)$ was estimated numerically with the MCMC method and a Metropolis-Hastings algorithm (see e.g. [35]). We refer the reader to Text S3 for details about the algorithm and the estimation process, and for more explicit results.

We analysed the posterior distribution for the replicate transmissibility $\psi_{\text{site}}(T, r)$ for each replicate r of treatment T. Figure 5A shows as an example the distributions for four different replicates of the same treatment. The mean of the posterior for $\psi_{\text{site}}(T, r)$ was chosen as the new estimated value of the transmissibility, $\hat{\psi}_{\text{site}}(T, r)$ (with error bar corresponding to the 95% credible interval, Figure 5B). The new estimates for the mean and variance for each population, $\langle\hat{\psi}\rangle_{\text{pop}}(T, r)$ and $\hat{\sigma}^2_{\text{pop}}(T, r)$ (with the associated confidence intervals), were re-calculated from Equation 0 using $\hat{\psi}_{\text{site}}(T, r)$, and plotted in the plane $\left(\langle\psi\rangle_{\text{pop}}, \sigma^2_{\text{pop}}\right)$ (Figure 5B). In order to assess the dependence of P_{inv} on the estimated values of $\langle\psi\rangle_{\text{pop}}$ and σ^2_{pop}, we used multiple logistic regression [41], supplemented by linear discriminant analysis (LDA) [42], to find the line in the plane $\left(\langle\psi\rangle_{\text{pop}}, \sigma^2_{\text{pop}}\right)$ that best separates the two groups of invasive and non-invasive replicates. All the tests were performed with the R statistical package [43].

Supporting Information

Table S1 Notional treatments of the population experiment. Parameter estimates used for the experimental design are shown here, compared with the corresponding *post-hoc* estimates from the population experiment. The treatments were devised to achieve an approximately constant value of $\langle\psi\rangle_{\text{pop}} \simeq 0.5$, and values of σ^2_{pop} decreasing by approximately regular intervals from 0 to 0.25. The total number of sites (with and without nutrient) for each population is $N = 217$. Columns 4 to 7: estimates of ψ_{site}, $\langle\psi\rangle_{\text{pop}}$, and σ^2_{pop} as a function of the nutrient concentration from pair experiments (see Text S2). For each parameter, the best-fit value is indicated in bold face, the 95% confidence interval is in parentheses. The suffix (*a*) indicates parameters obtained by interpolation between those for 3% and 5% nutrient concentrations. P_{inv} is calculated with numerical simulations (cf. manuscript Figure 1 and see Text S1 for details). Columns 8 to 11: summary statistics for the estimates $\hat{\psi}_{\text{site}}(T, r)$, $\hat{\psi}_{\text{pop}}(T, r)$ and $\hat{\sigma}^2_{\text{pop}}(T, r)$, in the form (mean $\pm$ standard deviation); distributions for treatments E and F are significantly asymmetric, see comment in Text S3. The estimate for the probability of invasion, $\hat{P}_{\text{inv}}$, is calculated as the ratio of the number of invading replicates divided by total number of replicates, for each treatment (cf. manuscript Figure 2). (PDF)

Text S1 We give a theoretical background on percolation theory and epidemic processes in heterogeneous systems. (PDF)

Text S2 We describe the colonisation experiments performed to find the value of the transmissibility as a function of the nutrient concentration. (PDF)

Text S3 We give more details on the MCMC methods used for parameter estimation. We also provide a complete survey of the results of the estimation, and show how heterogeneity in transmissibility can be described by a simple hierarchical model. (PDF)

Author Contributions

Conceived and designed the experiments: FMN FJP SNT WO CAG. Performed the experiments: AB WSF FMN. Analyzed the data: FMN. Wrote the paper: FMN FJP SNT WO CAG DJB.

References

1. Jules ES, Kauffman MJ, Ritts WD, Carroll AL (2002) Spread of an invasive pathogen over a variable landscape: a nonnative root rot on Port Orfort cedar. Ecology 83: 3167–3181.
2. Kauffman MJ, Jules ES (2006) Heterogeneity shapes invasion: host size and environment influence susceptibility to a nonnative pathogen. Ecol Appl 16: 166–175.
3. Segal S, Hill AVS (2003) Genetic susceptibility to infectious diseases. Trends Microbiol 11: 445–448.
4. Fryer HR, Frater J, Duda A, Roberts MG (2010) Modelling the evolution and spread of hiv immune escape mutants. PLoS Pathog 6: e1001196.
5. Ennos R, McConnell KC (2003) Variation in host resistance and pathogen selective value in the interaction between *Pinus sylvestris* and the fungus *Crumenulopsis sororia*. Heredity 91: 193–201.
6. Woolhouse MEJ, Dye C, Etard JF, Smith T, Charlwood JD, et al. (1997) Heterogeneities in the transmission of infectious agents: implications for the design of control programs. Proc Natl Acad Sci USA 94: 338–342.
7. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM (2005) Superspreading and the effect of individual variation on disease emergence. Nature 438: 355–359.
8. Matthews L, McKendrick IJ, Ternent H, Gunn GJ, Synge B, et al. (2006) Supershedding cattle and the transmission dynamics of *Escherichia coli* O157. Epidemiol Infect 134: 131–142.
9. Dwyer G, Elkinton JS, Buonaccorsi JP (1997) Host heterogeneity in susceptibility and disease dynamics: tests of a mathematical model. Am Nat 150: 685–707.
10. Dwyer G, Dusho J, ELkinton JS, Levin SA (2000) Pathogen-driven outbreaks in forest defoliators revisited: building models from experimental data. Am Nat 156: 105–120.
11. Smith DL, McKenzie FE, Snow RW, Hay SI (2007) Revisiting the basic reproductive number for malaria and its implications for malaria control. PLoS Biol 5: e42.
12. Matthews L, Low JC, Gally DL, Pearce MC, Mellor DJ, et al. (2006) Heterogeneous shedding of *Escherichia coli* O157 in cattle and its implications for control. Proc Natl Acad Sci USA 103: 547–552.
13. Diekmann O, Heesterbeek JAP (2000) Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis, and Interpretation. John Wiley.
14. Cook AR, Otten W, Marion G, Gibson GJ, Gilligan CA (2007) Estimation of multiple transmission rates for epidemics in heterogeneous populations. Proc Natl Acad Sci USA 104: 20392–20397.
15. Jeger MJ (1986) Asymptotic-behavior and threshold criteria in model-plant disease epidemics. Plant Pathology 35: 355–361.
16. Ball F (1986) A unified approach to the distribution of total size and total area under the trajectory of infectives in epidemic models. Adv Appl Prob 18: 289–310.
17. Diekmann O, Heesterbeek JAP, Metz JAJ (1990) On the definition and the computation of the basic reproduction ratio r o in models for infectious diseases in heterogeneous populations. J Math Biol 28: 365–382.
18. Addy CL, Longini IM, Haber M (1991) A generalized stochastic model for the analysis of infectious disease final size data. Biometrics 47: 961–974.
19. Dushoff J, Levin S (1994) The effects of population heterogeneity on disease invasion. Math Biosc 128: 25–40.
20. Lloyd AL, Zhang J, Root AM (2007) Stochasticity and heterogeneity in host-vector models. J R Soc Interface 4: 851–863.
21. Glass K, Kappey J, Grenfell BT (2004) The effect of heterogeneity in measles vaccination on population immunity. Epidemiol Infect 132: 675–683.
22. Cross PC, Johnson PLF, Lloyd-Smith JO, Getz WM (2007) Utility of r_0 as a predictor of disease invasion in structured populations. J R Soc Interface 4: 315–324.
23. Newman MEJ (2002) Spread of epidemic disease on networks. Phys Rev E 66: 016128.

24. Kenah E, Robins JM (2007) Second look at the spread of epidemics on networks. Phys Rev E 76: 036113.

25. Miller JC (2007) Epidemic size and probability in populations with heterogeneous infectivity and susceptibility. Phys Rev E 76: 010101(R).

26. Neri FM, Pérez-Reche FJ, Taraskin SN, Gilligan CA (2011) Heterogeneity in SIR epidemics on lattices. J R Soc Interface 8: 201–209.

27. Bailey DJ, Otten W, Gilligan CA (2000) Saprotrophic invasion by the soil-borne fungal plant pathogen *rhizoctonia solani* and percolation thresholds. New Phytol 146: 535–544.

28. Otten W, Bailey DJ, Gilligan CA (2004) Empirical evidence of spatial thresholds to control invasion of fungal parasites and saprotrophs. New Phytol 163: 125–132.

29. Otten W, Gilligan CA (2006) Soil structure and soil-borne diseases: using epidemiological concepts to scale from fungal spread to plant epidemics. Eur J Soil Sci 57: 26–37.

30. Gilligan CA, Truscott JE, Stacey AJ (2007) Impact of scale on the effectiveness of disease control strategies for epidemics with cryptic infection in a dynamical landscape: an example for a crop disease. J R Soc Interface 4: 925–934.

31. Davis S, Trapman P, Leirs H, Begon M, Heesterbeek J (2008) The abundance threshold for plague as a critical percolation phenomenon. Nature 454: 634–637.

32. Salkeld DJ, Salathé M, Stapp P, Jones JH (2010) Plague outbreaks in prairie dog populations explained by percolation thresholds of alternate host abundance. Proc Natl Acad Sci USA 107: 14247–14250.

33. Stauffer D, Aharony A (1991) Introduction to Percolation Theory. London: Taylor and Francis.

34. Grassberger P (1983) On the critical behavior of the general epidemic process and dynamical percolation. Math Biosci 63: 157–172.

35. Gelman A, Carlin JB, Stern HS, Rubin DB (2003) Bayesian Data Analysis. New York: Chapman & Hall/CRC.

36. Olsson S (1995) Mycelial density profiles of fungi on heterogeneous media and their interpretation in terms of nutrient reallocation patterns. Mycol Res 99: 143–153.

37. Jacobs H, Boswell GP, Scrimgeour CM, Davidson FA, Gadd GM, et al. (2004) Translocation of carbon by *Rhizoctonia solani* in nutritionally-heterogeneous microcosms. Mycol Res 108: 453–462.

38. Lloyd-Smith JO, Schreiber SJ, Getz WM (2006) Moving beyond averages: individual-level variation in disease transmission. Contemp Math 410: 235–258.

39. Kuulasmaa K (1982) The spatial general epidemic and locally dependent random graphs. J Appl Prob 19: 745–758.

40. Cox DR, Isham V (1980) Point Processes, volume 12 of *Monographs on Applied Probability and Statistics*. London: Chapman and Hall.

41. Forthofer RN, Lee ES, Hernandez M (2007) Biostatistics: a guide to design, analysis and discovery, 2nd edition. Academic Press.

42. McLachlan GJ (1992) Discriminant analysis and statistical pattern recognition. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. New York: John Wiley & Sons.

43. Dalgaard P (2008) Introductory Statistics with R, 2nd edition. New York: Springer.