Research article

The spread of incompatibility-inducing parasites in sub-divided host populations

Max Reuter*[†1], Laurent Lehmann[†2,4] and Frédéric Guillaume[†3]

Address: [1]Research Department for Genetics, Evolution and Environment, Faculty of Life Sciences, University College London, Wolfson House, 4 Stephenson Way, London, NW1 2HE, UK, [2]Department of Genetics, Cambridge University, Downing Street, Cambridge, CB2 3EH, UK, [3]Department of Zoology, University of British Columbia, 6270 University Blvd., Vancouver, BC V6T 1Z4, Canada and [4]Department of Biological Sciences, Stanford University, Stanford, CA, USA

Email: Max Reuter* - m.reuter@ucl.ac.uk; Laurent Lehmann - lehmann@stanford.edu; Frédéric Guillaume - guillaum@zoology.ubc.ca

* Corresponding author   †Equal contributors

Abstract

Background: Maternally transmitted symbionts have evolved a variety of ways to promote their spread through host populations. One strategy is to hamper the reproduction of uninfected females by a mechanism called cytoplasmic incompatibility (CI). CI occurs in crosses between infected males and uninfected females and leads to partial to near-complete infertility. CI-infections are under positive frequency-dependent selection and require genetic drift to overcome the range of low frequencies where they are counter-selected. Given the importance of drift, population sub-division would be expected to facilitate the spread of CI. Nevertheless, a previous model concluded that variance in infection between competing groups of breeding individuals impedes the spread of CI.

Results: In this paper we derive a model on the spread of CI-infections in populations composed of demes linked by restricted migration. Our model shows that population sub-division facilitates the invasion of CI. While host philopatry (low migration) favours the spread of infection, deme size has a non-monotonous effect, with CI-invasion being most likely at intermediate deme size. Individual-based simulations confirm these predictions and show that high levels of local drift speed up invasion but prevent high levels of prevalence across the entire population. Additional simulations with sex-specific migration rates further show that low migration rates of both sexes are required to facilitate the spread of CI.

Conclusion: Our analyses show that population structure facilitates the invasion of CI-infections. Since some level of sub-division is likely to occur in most natural populations, our results help to explain the high incidence of CI-infections across species of arthropods. Furthermore, our work has important implications for the use of CI-systems in order to genetically modify natural populations of disease vectors.

Background

Vertically transmitted symbionts are common in nature. They are of diverse phylogenetic origin, comprising viruses, bacteria, protozoa, and fungi, and reside in an equally diverse range of hosts, including both animals and plants[1-5]. Many vertically transmitted symbionts live inside the host cytoplasm and accordingly their transmission is strictly maternal. For these symbionts to spread through a host population, the number of infected daughters produced by an infected female must exceed the number of daughters produced on average in the population. Symbionts have evolved a number of strategies to achieve this goal. Some, such as secondary symbionts of aphids, are beneficial and increase the host's overall fitness through a positive effect on fecundity or survival [6,7]. Others spread by biasing the sex ratio of their host towards females. This strategy is common in symbionts of arthropods, which achieve sex ratio distortion by inducing parthenogenesis, feminizing genetic males, or selectively eliminating male hosts [2,8]. A third strategy differs fundamentally from the previous two in that it promotes infection indirectly by impeding the reproduction of uninfected female hosts rather than affecting that of infected ones [8-10]. This is achieved through a mechanism called cytoplasmic incompatibility (CI) which appears to have evolved at least twice independently in two bacterial symbionts of arthropods, *Wolbachia* [11] and *Cardinium* [12]. Incompatibility is presumably caused by a modification in the sperm of infected males that causes increased zygote mortality unless rescued by a symbiont present in the egg [2]. CI therefore specifically targets the fecundity of uninfected hosts.

Cytoplasmic incompatibility benefits infected individuals by reducing the fecundity of uninfected females in a targeted way. This reduction increases the relative fitness of infected females, i.e., their fecundity measured relative to the average fecundity in the population. The degree to which CI increases the relative fitness of infected depends on the frequency of incompatible matings, making selection on CI frequency-dependent [9,13-16]. Whether CI is selectively favored depends on the balance between the increase in relative fitness of infecteds due to CI and forces that act against its spread, namely a fecundity cost of bearing an infection and imperfect transmission of the symbiont. In the presence of a cost of infection and/or imperfect transmission, an unstable infection equilibrium exists (often referred to as the 'invasion threshold'), separating the lower frequency range in which infection is counter-selected and the higher frequency range where it is selected for.

For symbionts to spread in a host population, infection frequency has to exceed the unstable equilibrium. Many authors have therefore made the verbal argument that random genetic drift would be required for the successful spread of CI-symbionts [9,17]. Egas et al. [17] tested this conjecture in computer simulations that determined the probability of CI spread in random mating populations of different sizes (and hence subject to varying levels of genetic drift). The authors concluded that random drift was unlikely to drive infection frequency beyond the invasion threshold and allow it to go to fixation. This result is at odds with the high number of species that have been found to be infected with CI-inducing parasites [see [18] for a review]. Egas et al. therefore conclude that in addition to CI, infection spread is driven by factors such as other reproductive manipulations, effects of infection on the population sex ratio or other fitness-compensating effects [17].

The conclusion of Egas et al.'s seems pessimistic for the spread of CI infection in panmictic populations of large size. However, their data also show that CI can spread with considerable probabilities when populations are small. This raises the possibility that the spread of CI might be facilitated if a large population is subdivided into a number of discrete demes that are linked by migration. Population sub-division is known to induce increased levels of local genetic drift, which could drive infection frequency beyond the invasion threshold within demes. Despite its clear relevance, the effect of genetic drift on CI spread in sub-divided population has found little treatment in the theoretical literature. Some studies have analyzed CI-dynamics either in continuous [19] or discrete space [20], but have done so using deterministic models which ignore the stochastic effect due to drift. The only model including drift (and only implicitly) has come to the surprising prediction that population structure would impede the spread of infection. In this model, Wade and Stevens [21] assume that hosts aggregate in groups to mate and reproduce before being mixed and compete at the scale of the whole metapopulation. Frequency change of CI was then calculated as a function of the variance in infection frequency between breeding groups. Wade and Stevens showed that increasing this variance, as would happen because of drift or if infection appeared in a subset of the groups, raised the invasion threshold and reduced the per-generation change in average infection frequency across groups.

Motivated by the discrepancy between biological intuition and theoretical prediction, we present here a new analysis of CI dynamics in sub-divided host populations. We first derive an analytical model which illustrates the different components of the selection pressure acting on a CI phenotype and how they are affected by population sub-division. This model allows us to derive the conditions under which a rare infection can spread in a population and to determine the expected equilibrium infection

frequency. We complement the analytical model with individual-based simulations. These relax some of the simplifying assumptions made in the model and generate predictions for more complex population structures.

Results
Analytical model
Similar to previous models treating the panmictic case we will ignore the dynamics of symbiont populations within individual hosts and concentrate on the presence or absence of parasite infection. This allows us to conveniently model the spread of infection as a two state model in which individuals are in one of the two states 'infected' and 'uninfected'. Transmission of infection state is purely maternal, resembling symbiont transmission in natural host populations. Vertical transmission of infection is assumed to be perfect and all offspring of infected mothers are infected. This assumption is justified because natural parasites causing CI usually show high rates of vertical transmission [9], but is relaxed in simulations presented below.

We assume that hosts live in a population composed of an infinite number of demes, each containing a finite number N of breeding males and N females (i.e., infinite island model assumptions). Hosts are diploid and semelparous (no overlapping generations) and undergo the following lifecycle: (1) Adult hosts mate randomly within demes. (2) Adult hosts reproduce. Each female produces a large number of offspring that is a function of the infection status of her and her mating partner. The fecundity of an uninfected female mated to an infected male (incompatible mating), relative to that of uninfected female mated to an uninfected male, is $(1 - B)$ and the relative fecundity of an infected female is $(1 - C)$, independently of the genotype of her mate. In these fecundities, C is the fecundity cost of infection borne by infected females and B is the fecundity cost of cytoplasmic incompatibility borne by uninfected females mated to infected males (note that for clarity we have chosen to deviate from the usual notation in which $C = s_f$ and $B = s_h$). All adults die after reproduction. (3) Juvenile hosts disperse. Male and female juveniles disperse with equal probabilities and randomly, meaning that any non-natal deme is reached with same probability. The probability that a juvenile is sampled in its natal deme after dispersal is given by $(1 - m)$ where m is the migration rate. (4) Juveniles compete for access to reproduction. Competition is sex-specific and occurs within demes. In each deme, N juveniles of each sex reach adulthood, the remainder die. This life cycle drastically differs from that of Wade & Stevens [19], where mixing of individuals is complete after breeding and occurs at the scale of the whole population. However, our model converges towards that of Wade & Stevens at the higher migration limit $m \rightarrow 1$.

Assuming this life-cycle and weak selection (small C and B), we show in the Appendix that the change in frequency (Δp) of infection in the population can be written as

$$\Delta p = p(1-p)\left[-C(1 - F_{st}) + B(1-m)^2\left(F_{st} - F_3^R \right) + pB\left(1 - g(N,m) \right) \right]$$

(1)

where F_{st}, F_3^R, and $g(N, m)$ are terms describing the effect of population sub-division on the probability that individuals in a deme share their infection status. More precisely, F_{st} is Wright's measure of population structure describing the probability that two randomly sampled individuals within a deme share their infection status by descent, F_3^R is the probability that one male and two females sampled with replacement from the same deme share their infection status by descent, and $g(N, m)$ is a function depending on deme size and migration rate (see eq. 21 in the Appendix).

The selective pressure on infection (the term in square brackets) comprises three components. The first, $-C(1-F_{st})$, describes the effect of directional selection against infection due to the fecundity cost C borne by infected females. The fecundity cost is weighted by the probability $(1-F_{st})$. This weighting expresses the fact that population sub-division reduces the strength of this component of selection. Specifically, genetic drift leads to increased coalescence within demes, thereby homogenising local demes with respect to infection status and levelling differences in fecundity between competitors. The second and third components of selection capture the change in infection frequency due to the benefit of CI. These arise whenever an infected female enjoys increased relative fitness due to the reduced productivity of an uninfected females having an incompatible mating with an infected male. Accordingly, the benefit of CI is proportional not only to the cost of incompatibility, B, but also to the frequency with which one infected female, one infected male and one uninfected female occur in a same deme (the first being the beneficiary, the latter two the incompatible mates). Equation 1 distinguishes two cases that can lead to this constellation, based on the genetic relationship between the infected female and male. The first of the benefit terms, $B(1-m)^2(F_{st} - F_3^R)$, is frequency-independent (does not contain p) and measures the beneficial effect arising from events where the infected male and female are related, i.e. are both infected due to recent coalescence. Accordingly, the term weights the cost of incompatibility, B, by the

term $(F_{ST} - F_3^R)$ which quantifies the joint probability that within a deme, a female shares her infection status by recent descent with an infected male mated to an uninfected female (who will suffer from incompatibility). Local genetic drift affects the term $(F_{ST} - F_3^R)$ in a non-trivial way, because on one side it increases the probability that a female shares the same infection status as a male, but on the other side it decreases the probability that this male will have mated with a female bearing the alternative status. As a consequence of these opposing effects, $(F_{ST} - F_3^R)$ is not necessarily a monotonic function of N and m. The weighting by $(1 - m)^2$ expresses the fact that the off-spring of an infected female have to remain in their natal patch and compete against other sedentary offspring in order for an increase in the relative fitness of infected females to arise. The second benefit term, $pB(1 - g(N, m))$, measures the benefit of CI that is independent of recent coalescence. Here, the infected male and female are unrelated in the sense there is no recent coalescence in their infection status. Accordingly, the probability of such an event is frequency-dependent and the term includes the average frequency p of infection in the population. The term is further offset by a quantity $g(N, m)$ which measures the effect of population sub-division on the frequency with which such a benefit occurs. The function $g(N, m)$ decreases monotonically in N and m (see Appendix), reflecting the fact that with reduced population sub-division, the benefit of CI relies more and more on interactions between unrelated individuals.
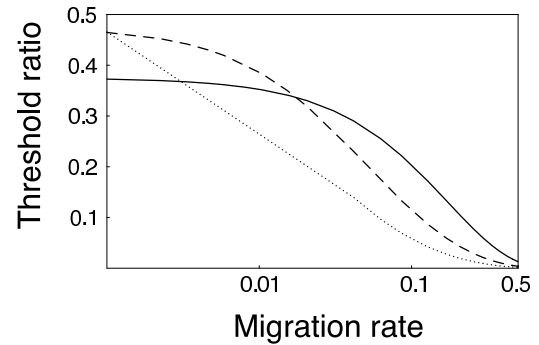
Equation 1 also allows us to recover the dynamics of CI infections considered by previous treatments. When deme size becomes very large $(N \rightarrow \infty)$ or when migration is complete $(m \rightarrow 1)$, equation 1 reduces to

$$\Delta p = p(1 - p) [-C + pB], \qquad (2)$$

which is the weak-selection equivalent of the frequency change calculated by [14].

We will now consider the situation when infection is rare $(p \rightarrow 0)$. This situation is highly relevant, because it is the natural condition faced by all invading CI-infections and the dynamics of infection at this point determine the probability of invasion (or, in mathematical terms, the stability of the equilibrium $p = 0$). Taking the lower frequency limit of equation 1, we can derive the condition for the invasion of a rare CI-infection in terms of the cost-to-benefit ratio C/B,
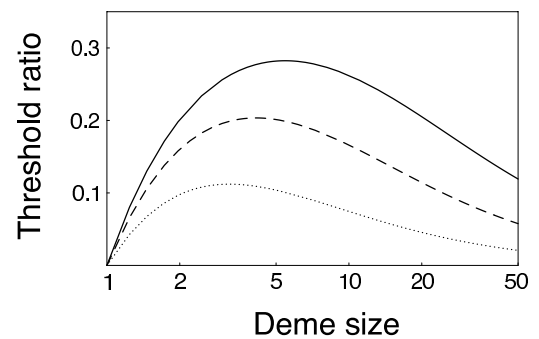

Figure 1
Analytical predictions on the effect of population structure on CI-dynamics. The figure shows the threshold cost-to-benefit ratio (C/B) for the invasion of CI-symbionts. A. Threshold ratio as a function of host migration rate for deme sizes of $N = 4$ (solid line), $N = 20$ (dashed line), and $N = 50$ (dotted line). B. Threshold ratio as a function of deme size for host migration rates of $m = 0.05$ (solid line), $m = 0.1$ (dashed line), and $m = 0.2$ (dotted line).

$$\frac{C}{B} < (1-m)^2 \frac{(F_{st} - F_3^R)}{(1 - F_{st})}. \qquad (3)$$

The expression on the right-hand size of this inequality is a function of deme size (N) and migration rate (m). Equating both sides of inequality 3, we can define a threshold cost-to-benefit ratio above which infections will not be able to invade the population. Figure 1 shows how this threshold ratio varies with both host migration rate and deme size. For population with very large deme sizes and very high migration rates (i.e., quasi-panmixia) the threshold approaches zero and infections cannot invade unless they are virtually cost-free relative to the intensity of incompatibility. With decreasing migration rate, however, the threshold ratio increases monotonically, indicat-
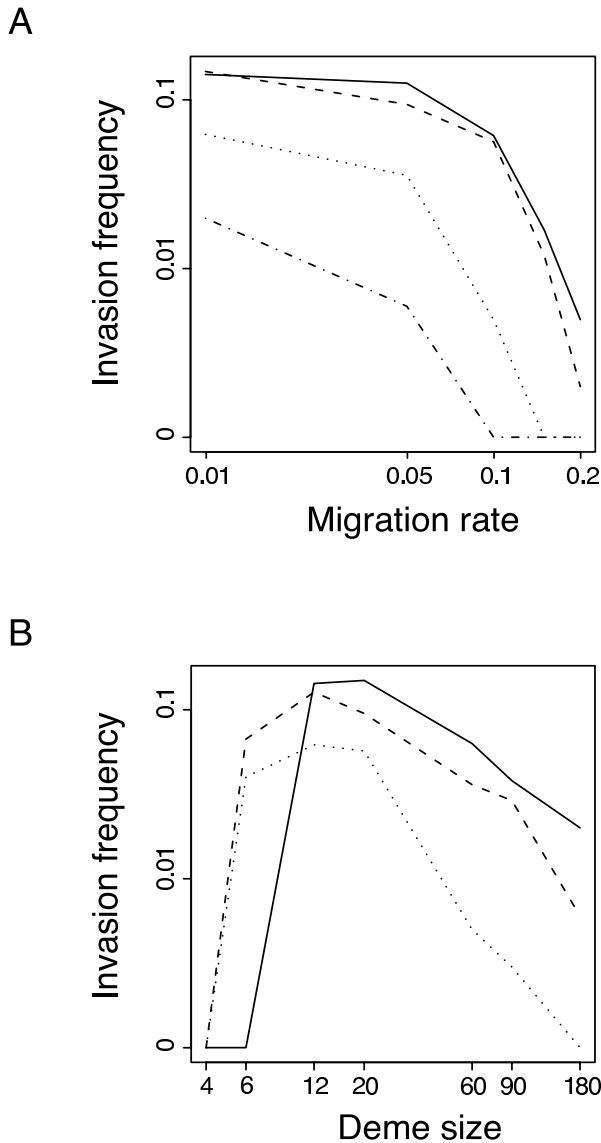
A

B

Figure 2
Simulation results on the effect of population structure on CI-invasion frequency. The figure shows A, the frequency of invasion as a function of migration rate for deme sizes of N = 12 (solid line), N = 20 (dashed line), N = 60 (dotted line) and N = 180 (dashed-dotted line) and B, the frequency of invasion as a function of deme size for host migration rates of *m* = 0.01 (solid line), *m* = 0.05 (dashed line), and *m* = 0.1 (dotted line).

ing that infections can spread even if associated with a fecundity cost (C>0) and/or incomplete incompatibility (B<1). The change of the threshold ratio with deme size is non-monotonous, the threshold ratio being highest at intermediate deme sizes. This effect can be interpreted in terms of the impact of local drift on the frequency of

incompatible matings. In host populations with intermediate deme size, drift maintains the local variance in infection at a level that maximizes the benefit of incompatibility, making such population most prone to the invasion of CI-infections.

Simulations

To complement the analytical model, we ran individual-based computer simulations of CI spread, based on an extended version of the population genetics software Nemo [22]. The simulations allowed us to relax some of the more idealistic assumptions of the model and measure additional parameters of infection dynamics, such as the speed of spread and the equilibrium frequency attained.

The simulations considered a finite population of 1800 individuals (half males, half females) subdivided into of n_d demes of *N* individuals each. The demes are linked by migration at rate *m*. We implemented two migration models, an island model (all demes equidistant) and a stepping stone model (populations arranged on a ring). Furthermore, in one set of simulations we explored the effect of differential male and female migration in an island model.

The life cycle is identical to that assumed in the analytical model. Male and female fecundity was assumed Poisson distributed and each individual was assigned a number of offspring drawn form a distribution with mean and variance *f* = 10. The cost of infection was implemented by reducing the fecundity of infected females by a factor (1 - *C*). Cytoplasmic incompatibility acted upon fertilization. An uninfected egg fertilized by the sperm of an infected father died with probability *B*. In all simulations, we relaxed the assumption of perfect vertical transmission made in the analytical model. Thus, we assumed that infection was lost with a probability μ from mother to offspring.

Throughout this study we used a standard set of infection parameters with an incompatibility cost of *B* = 0.5, a fecundity cost of *C* = 0.01, and a rate of parasite loss during vertical transmission of μ = 0.05. Based on the model for an infinite panmictic population [14], an infection with these parameters is selectively favoured when its frequency exceeds the invasion threshold of p_{it} = 0.13 and will attain an equilibrium frequency of p_e = 0.94. The parameter values used are inspired by empirical data obtained in populations of *Drosophila simulans* harbouring a CI-Wolbachia [23]. Additional simulations (data not shown) showed that varying the values has the effect expected under the deterministic model of infection dynamics with panmixia.

At the start of each simulation, one deme was inoculated with one infected female and one infected male. This starting frequency is well below the invasion threshold of p_{it} = 0.13 (and hence counter-selected) when considering the whole population of 900 males and 900 females, as well as when considering deme sizes that exceed 7 males and 7 females. Simulations were run for 2000 generations and infection frequency was recorded for each deme at intervals of 10 generations. We considered an infection to have invaded if it was still present at the end of the simulation run (generation 2000). Simulations were replicated 1000 times for any combination of parameters.

Effects of dispersal rate and deme size in an island model
In a first set of simulations, we analyzed the dynamics of infection as a function of host migration rate and deme size in an island model of population structure. In these simulations, the population was sub-divided into demes of sizes N = {4, 6, 12, 20, 60, 90, 180}. Since the sex ratio was assumed to be even in our simulations, these deme sizes are equivalent to 2, 3, 6, 10, 30, 45, and 90 reproductive females. With a total population size of 1800, the array of deme sizes imply deme numbers of n_d = {450, 300, 150, 90, 30, 20, 10}. Demes were linked by male and female migration at equal rates. The range of migration rates used was m = {0.01, 0.05, 0.1, 0.2}. We ran simulations for all combinations of migration rate m and deme size N.

Frequencies of invasion (i.e., the proportions of replicate simulation runs in which the infection was present after 2000 generations) varied with both deme size and migration in a way that fit the predictions of our analytical model (Fig. 2). While increasing host migration rate lead to a decline in invasion frequency (Fig. 2A), infections were more likely to persist when deme size was intermediate (Fig. 2B). Figure 2 shows that population sub-division (restricted migration) allows CI infection to invade populations it would not have invaded under panmixia. For example, with deme sizes of N = 180, infection can sometimes invade with migration rates of m = 0.01 and m = 0.05, but not with m = 0.1 and above, which for ten demes corresponds to panmixia.

In addition to its effect on the probability of spread, population sub-division also influenced the equilibrium frequency they attained. Generally, infections attained a higher equilibrium frequency (prevalence) under conditions that minimized the amount of local genetic drift. Thus, prevalence in populations with strong drift (low migration rate and/or small deme size) remained considerably lower than the value predicted for panmixia (p_e = 0.94) at which populations with weak drift (large demes and/or high rates of migration) stabilized (Fig. 3). The low prevalence in high-drift populations is not due to a failure
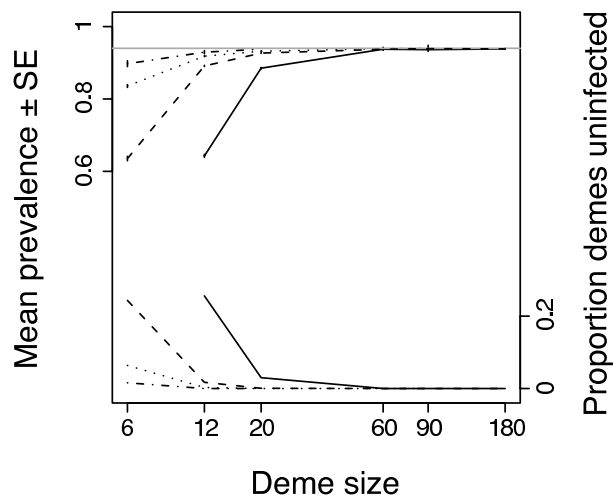

Figure 3
Simulation results on the mean population prevalence and between-deme polymorphism in infection.
Mean population prevalence (± SE) and the proportion of demes infected as a function of deme size in simulations with equal migration rate in males and females. Results are shown for host migration rates of m = 0.01 (solid line), m = 0.05 (dashed line), m = 0.1 (dotted line), and m = 0.2 (dashed-dotted line). The grey line indicates the equilibrium frequency of p_e = 0.94 predicted for a panmictic population of infinite size.

to attain equilibrium. Rather, low prevalence is the result of two mechanisms. First, drift causes prevalence to fluctuate around the equilibrium of p_e = 0.94. Because infection frequency cannot exceed 100%, large stochastic fluctuations in populations with strong drift will tend to decrease mean prevalence. Second, under conditions of strong local drift, population-wide mean prevalence is lowered by the total loss of infection in some demes (Fig. 3). For instance, in the two cases showing the lowest mean prevalence (N = 12 with m = 0.01, and N = 6 with m = 0.05), around a quarter of demes had lost infection while the prevalence in the remaining infected demes was 86–88% (compared to a population-wide prevalence of about 64%). Additional simulations have shown that such polymorphism is stable over large time-spans (more than 5,000 generations).

Effects of dispersal rate and deme size in a stepping stone model
The island model used in the analytical model and the above simulations is convenient from a modelling point of view, but is representative of few, if any, natural situations. Indeed, it assumes that all demes of a population are equidistant, while in natural populations, demes tend to be arranged in a geographic pattern that causes some pairs of demes to be further apart than others. Accord-

ingly, some demes are more likely to exchange migrants than others. In order to assess the influence of geography on the spread of CI, we performed a set of simulations in a simple stepping stone setting, in which the demes were arranged on a circle. Migration occurred between neighbouring demes along the circle. Simulations were run for deme sizes of N = {4, 6, 12, 20, 60, 180} (deme numbers n_d = {450, 300, 150, 90, 30, 10}) and migration rates of m = {0.01, 0.05, 0.1, 0.2, 0.4, 0.6}.

Figure 4 shows the frequency of invasion observed in these simulations as a function of migration rate (4A) and deme size (4B). Overall, the results are qualitatively equivalent to those obtained under the island model. Thus, the frequency of CI-invasion tends to decrease with increasing host migration rate. Furthermore, invasion frequency is highest for intermediate deme size. However, comparing Figures 2 and 4 shows that there are quantitative differences between the island model and the stepping stone model. In particular, the explicit geographical arrangement of demes in the stepping stone model appears to increase the probability of CI spread with high rates of host migration. This difference could arise because the limiting migration to adjacent demes in the stepping stone model reduces the degree to which infection is diluted across the entire population, making it more likely that the infection frequency will remain above the infection threshold in groups of neighbouring demes.

Infection dynamics as a function of male and female dispersal rates
In a third set of simulations, we addressed the relative importance of male and female dispersal rates for the spread of CI infections. We ran simulations for all combinations of male and females migration rates of m = {0.01, 0.05, 0.1, 0.2}. Population sub-division was fixed to n_d = 30 and N = 60.

The parasites being maternally transmitted, female dispersal was a major determinant of infection dynamics. As shown in Figure 5, the frequency of CI invasion decreased with female dispersal rate. However, male migration rate also had an effect, in that female philopatry favoured the spread of infection more when males dispersed at a low or moderate rate and less if male dispersed at a high rate.

Discussion
In this paper we have investigated the effect of host population structure on the spread of incompatibility-inducing symbionts. Our analyses demonstrate that population structure can greatly facilitate the spread of CI. In our simulations, infections spread from being present in a single individual of each sex (p_f = p_m = 1/900) to an equilibrium frequency of $p \approx 0.94$ with probabilities as high as 15% (cf. Figs. 2 and 4), despite the relatively benign level of incompatibility used (B = 0.5). These results suggest that
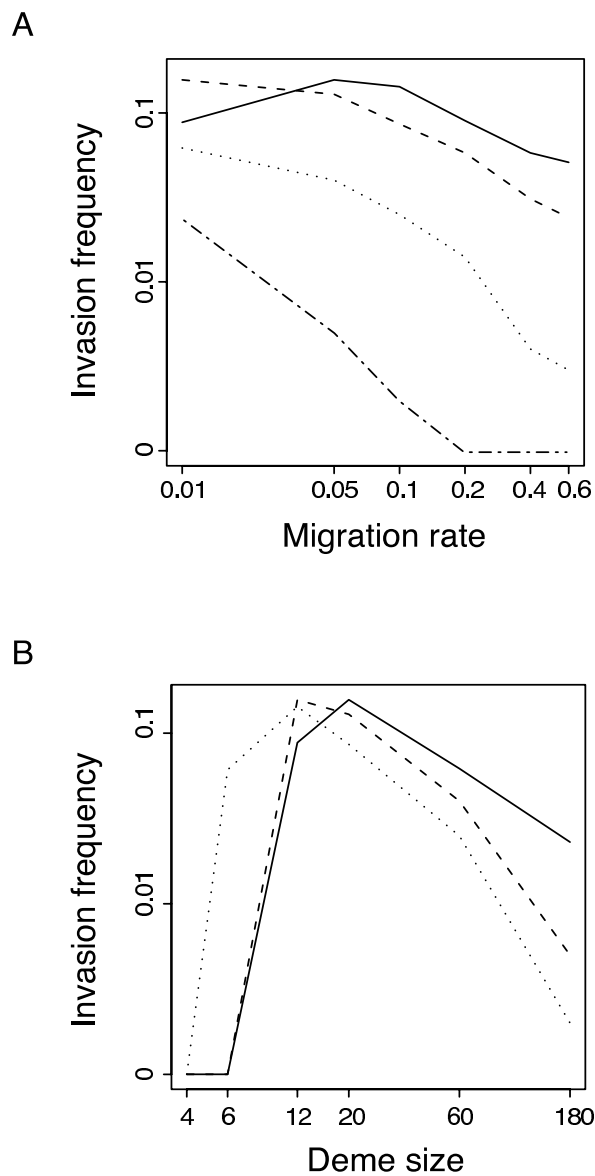

Figure 4
Simulation results on the effect of population structure on CI-invasion frequency in a stepping stone model. The figure shows the frequency of invasion, A, as a function of migration rate for deme sizes of N = 12 (solid line), N = 20 (dashed line), N = 60 (dotted line) and N = 180 (dashed-dotted line) and B, as a function of deme size for host migration rates of m = 0.01 (solid line), m = 0.05 (dashed line), and m = 0.1 (dotted line). See Figure 2 for the results with an island model.

the invasion threshold predicted by panmictic models does not represent a major barrier to invasion of CI into structured populations. If regular opportunities exist for horizontal transmission of infections into the popula-
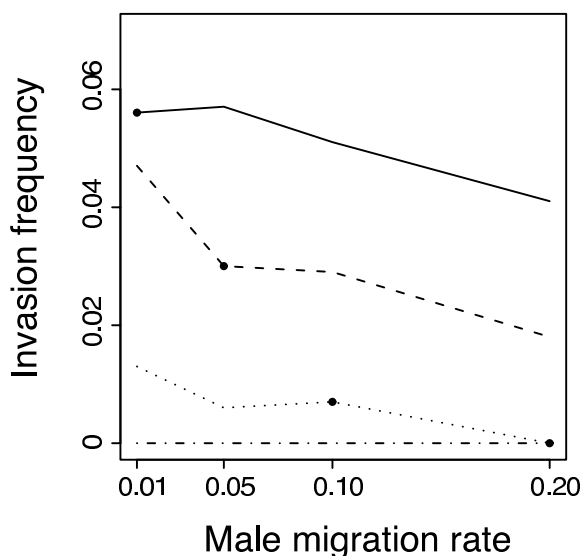
Figure 5
Simulation results on the effect of sex-specific dispersal rates on CI-invasion frequency. Invasion frequency is shown for different combinations of male migration rate (on the abscissa) and female migration rate (mf = 0.01: solid line; m_f = 0.05: dashed line; m_f = 0.1: dotted line; m_f = 0.2: dashed-dotted line). Dots indicate combinations of equal male and female migration.

tions, the spread of infection is only a matter of time. Population sub-division, which can be assumed to affect most natural populations to some degree, could thus provide the key to understanding the high incidence of CI-inducing symbiont infections that has been observed in arthropods, and in particular insects [18].

Both the model and the simulations consistently showed that the rate of invasion of CI-symbionts varies according to both parameters of host population structure considered in our study, deme size and migration rate. The latter of these parameters affects infection in a straightforward way; high levels of host migration will result in the dilution of infection out of a deme into the population at large and therefore decrease the probability of successful invasion. Furthermore, CI spread relies on limited dispersal of both males and females, in order to assure that infected females enjoy an increased relative fitness as a result of uninfected females suffering incompatible matings. The other parameter, deme size, shows an interesting non-monotonous effect whereby CI-infections invade easiest when deme size is intermediate. Under these conditions, local genetic drift creates heterogeneity in infection among the local breeding population that guarantees a high proportion of incompatible matings and hence a

large selective advantage for infection. Small deme size does not favour the spread of infection for two reasons. First, strong drift results in the frequent elimination of infection shortly after the initial inoculation of a population. Second, in the event that infection is not lost but fixed in one deme, it is subject to negative between-deme selection because the fecundity cost of infection reduces the productivity of infected demes. Large deme sizes, in turn, are disadvantageous because local drift is not strong enough to overcome the fitness disadvantage that CI-infections suffer at low frequencies due to the cost of infection.

As outlined above, the non-monotonous dependence of invasion probability on deme size results partly from selection against demes fixed for either the uninfected or the infected state. Infected demes lose out in between-deme competition because their productivity is reduced by the cost of infection. Accordingly, the low invasion success of a CI strategy in populations with small demes is related to the fact that infection reduces fecundity irrespectively of whether CI results in harm to uninfected individuals or not. Invasion success in small demes is expected to be higher for agents (be it genes or symbionts) in which the fecundity cost is conditional on the expression of the harming act (such as CI). This is the situation usually considered in models of the evolution of spiteful behaviour [24,25]. In these models, invasion of spiteful traits has been found to decrease monotonously with deme size.

The simulation results presented here demonstrate not only that population sub-division affects CI-dynamics, but also that this effect varies according to the population model considered. Thus, host migration had a stronger negative effect on CI invasion in the island model than in the stepping-stone model. This quantitative difference can be explained by a lesser degree of dilution of infection in the geographically explicit stepping stone model. Thus, local migration in this population setup resembles diffusion and is expected to result in a spatial correlation of infection frequency, with adjacent demes being more similar in infection than demes further apart. Accordingly, migration out of infected demes will drive up infection frequency in adjacent demes very effectively and is more likely to bring infection frequency in the range of positive selection for infection. The flipside of localized migration is, however, that the spread of infection across the population as a whole proceeds more slowly and more generations are needed to reach the equilibrium prevalence in the population as a whole (data not shown).

Our prediction that population sub-division can favour the spread of a CI infections differs diametrically from that of Wade and Stevens [21]. These authors found that

population structure impedes the spread of CI. A closer look at the two models reveals that this discrepancy is rooted in the different ways in which population structure is implemented in the life-cycle underlying the models. Wade and Stevens' model assumes that reproduction takes place within groups. Offspring produced in different matings groups then undergo complete dispersal and compete at the level of the entire population [hard selection sensu [26]], before being partitioned in groups again for the next round of reproduction. Population structure is implemented as variance in infection frequency between mating groups, which is a fraction of the total variance in infection frequency $p(1-p)$. Importantly, the portion of the total variance in infection that is attributed to the between-deme component is arbitrarily determined by a parameter. Accordingly, between- and within-deme variance are both assumed to vary completely independently of the migration rate and are thus independent of the geographical structure of the population. In this setting, increasing the variance in infection between mating groups will necessarily be disadvantageous for the spread of infection, because it is associated with a more homogeneous infection within demes and hence a reduced frequency of incompatible matings. The spread of CI is further compromised by the assumption that competition occurs at the level of the entire population. Global competition reduces the net advantage of CI because fitness is measured by comparing the fecundity of infected females to the population average rather than the deme average. As a consequence, a high frequency of incompatible matings within some individual demes will only generate a small increase in the relative fitness of infected females across the entire population. In our model, population structure is not imposed as an external parameter but arises naturally as a consequence of limited dispersal between permanent demes. Both reproduction and competition take place within demes, but separated by an episode of dispersal. With such a setting, intermediate between hard and soft selection, population structure affects both the frequency of incompatible matings and the probability that infected individuals will benefit from the increase in relative fitness brought about by CI. Our model shows that, if implemented in this natural way, population sub-division is conducive to the spread of CI-symbionts because the benefit of CI is likely to go to infected individuals rather than being diluted in the population at large. Under the assumption that competition occurs before dispersal (pure soft selection), we expect that population sub-division would be even more in favor of the spread of CI than under our life cycle because competition would occur only locally, thus increasing the benefits of reducing the fecundity of neighbouring uninfected females. This suggests that our results are qualitatively robust to changes of assumptions on the timing of dispersal.

The analytical model presented here assumes that individuals perform a large number of random matings within demes. Violations of this assumption due to systematic inbreeding or inbreeding avoidance alter the dynamics of an incompatibility-inducing mutant. Inbreeding reduces the expected number of incompatible matings and hence impedes the spread of CI. Outbreeding, on the contrary, increases the number of incompatible matings and promotes the spread of infection [27]. These effects of the mating system are expected to occur irrespective of population structure and hence add to the effects of dispersal rate and deme size predicted here. However, population structure is expected to have an influence on the quantitative difference between the frequency change with outbreeding/inbreeding and that with random mating. Both the advantage of outbreeding and the disadvantage of inbreeding are subject to the presence of diversity in infection within the mating populations. Under conditions of strong drift, the mating system can therefore be expected to have a small effect on infection dynamics.

In addition to a simple mating system, random mating, our model and simulations assumed non-overlapping generations. While this assumption might hold for some real-world cases, the majority of species show some degree of overlap between generations. Population age structure has been suggested to negatively affect the dynamics of CI and impeding infection spread [28], although the assumptions of this model are unclear. We have not extended our model in this direction but we would expect that including age structure would have a positive effect on the probability of CI spread. With overlapping generations, only a fraction of the breeding population is replaced at every generation. As a consequence, increasing age structure will, for a given deme size, reduce the part of the breeding population that has just reached adulthood and potentially migrated. Thus, age structure will tend to reduce the effective migration rate and accentuate population sub-division, thereby increasing the probability of CI spread (cf. Fig. 2A).

The predictions generated by our model and simulations have several implications for empirical research on CI-inducing symbionts. First, they generate the testable prediction that symbionts causing CI should be relatively more frequent in species or populations that show a higher degree of sub-division. Furthermore, our results on overall prevalence (Fig. 3) highlight potential problems for obtaining meaningful estimates of prevalence. Most natural hosts populations will be sub-divided and potentially subject to strong genetic drift. Accordingly, the effective population size was found to be around ten times smaller than the numerical population size in a large survey of data on wild animal populations [29]. Under conditions of local drift, measures of prevalence should be

based on screens for infection in several demes, because strong drift can generate local variation in infection frequency. A wide sampling screen is the more important the lower the migration rate of hosts and the smaller the size of local breeding populations.

Finally, our results are of fundamental importance with regards to the exploitation of CI as a drive-systems for genetic manipulation of natural populations of arthropods. Introducing transgenes into natural populations can be desirable for a number of reasons, the most important of which is the aim of controlling the transmission of vector-borne disease such as malaria. The realization that CI can bring about the rapid spread of symbiont infection in large host populations has led to the proposition of using the power of this system to drive transgenes through a host population [e.g., [30,31]]. However, previous analyses of the dynamics of such a system based on a panmictic model suggested that large numbers of genetically modified organisms would be required in order to overcome the invasion threshold to achieve successful spread of the construct [31]. Our results suggest that this limitation applies to a much lesser degree to situations in which vector populations are sub-divided. As shown in Figures 2 and 4, infections in highly subdivided populations spread with reasonable probability, even if populations are inoculated with a single pair released in one deme. The probability of successful invasion is expected to be much increased if the number of released individuals is large relative to the size of local demes. By reversing their logic, the predictions of our model further suggest that invasion success could be increased by artificially limiting host migration. A combined strategy of local release of transgenic hosts combined with, for instance, the pesticide treatment of adjacent demes would help the establishment of infection in the focal deme, creating a base for subsequent spread through the remainder of the population.

Methods
All methods are described in the main text. Additional information on the derivation of the analytical model is provided in the Appendix (Additional file 1).

Conclusion
Our work has shown that population sub-division greatly facilitates the spread of CI-inducing symbionts and allows infections to invade large populations, even when initially rare. Our study therefore helps to reconcile the high incidence of CI-infections across arthropod species with the considerable barriers to invasion predicted by models of CI-dynamics based on large panmictic populations. Furthermore, our predictions validate the use of CI-systems for the genetic manipulation of natural populations. Indeed, sub-divided populations of hosts could be suc-

cessfully infected with relatively small numbers of genetically engineered CI-agents, in particular if gene flow can be artificially restricted during the early periods of invasion.

Authors' contributions
MR conceived the research, MR, LL, and FG performed the research. MR, LL, and FG wrote the paper.

Additional material

Additional file 1
The Appendix contains a detailed derivation of the analytical model.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-8-134-S1.pdf]

References
1. Schardl CL, Leuchtmann A, Spiering MJ: **Symbioses of grasses with seedborne fungal endophytes.** *Annu Rev Plant Biol* 2004, **55:**315-340.
2. Werren JH: **Biology of Wolbachia.** *Ann Rev Entomol* 1997, **42:**587-609.
3. Ruby EG, McFall-Ngai MJ: **A squid that glows in the night - development of an animal- bacterial mutualism.** *J Bacteriol* 1992, **174(15):**4865-4870.
4. Stasiak K, Renault S, Federici BA, Bigot Y: **Characteristics of pathogenic and mutualistic relationships of ascoviruses in field populations of parasitoid wasps.** *J Insect Physiol* 2005, **51(2):**103-115.
5. Hogg JC, Ironside JE, Sharpe RG, Hatcher MJ, Smith JE, Dunn AM: **Infection of Gammarus duebeni populations by two vertically transmitted microsporidia; parasite detection and discrimination by PCR-RFLP.** *Parasitology* 2002, **125:**59-63.
6. Oliver KM, Russell JA, Moran NA, Hunter MS: **Facultative bacterial symbionts in aphids confer resistance to parasitic wasps.** *Proc Natl Acad Sci U S A* 2003, **100(4):**1803-1807.
7. Ferrari J, Darby AC, Daniell TJ, Godfray HCJ, Douglas AE: **Linking the bacterial community in pea aphids with host-plant use and natural enemy resistance.** *Ecol Entomol* 2004, **29(1):**60-65.
8. Veneti Z, Reuter M, Montenegro H, Hornett EA, Charlat S, Hurst GDD: **Interactions between inherited bacteria and their hosts: The Wolbachia paradigm.** In *The influence of cooperative bacteria on animal host biology* Edited by: McFall Ngai MF, Henderson B, Ruby NG. Cambridge , Cambridge University Press; 2005:119-142.
9. Hoffmann AA, Turelli M: **Cytoplasmic incompatibility in insects.** In *Influential Passengers* Edited by: O'Neill SL, Hoffmann AA, Werren JH. Oxford , Oxford University Press; 1997:42-80.
10. Rousset F, Raymond M: **Cytoplasmic incompatibility in insects: why sterilize females.** *Trends Ecol Evol* 1991, **6(2):**54-57.
11. Yen JH, Barr AR: **New hypothesis of the cause of cytoplasmic incompatibility in Culex pipiens L.** *Nature* 1971, **232:**657-658.

12. Hunter MS, Perlman SJ, Kelly SE: **A bacterial symbiont in the Bacteroidetes induces cytoplasmic incompatibility in the parasitoid wasp Encarsia pergandiella.** *Proc R Soc Lond Ser B-Biol Sci* 2003, **270(1529):**2185-2190.
13. Fine PEM: **On the dynamics of symbiote-dependent cytoplasmic incompatibility in Culicine mosquitoes.** *Journal of Invertebrate Pathology* 1978, **30:**10-18.
14. Hoffmann AA, Turelli M, Harshman LG: **Factors affecting the distribution of cytoplasmic incompatibility in Drosophila simulans.** *Genetics* 1990, **126(4):**933-948.
15. Caspari E, Watson GS: **On the evolutionary importance of cytoplasmic sterility in mosquitoes.** *Evolution* 1959, **13:**568-570.
16. Turelli M: **Evolution of incompatibility-inducing microbes and their hosts.** *Evolution* 1994, **48(5):**1500-1513.
17. Egas M, Vala F, Breeuwer JAJ: **On the evolution of cytoplasmic incompatibility in haplodiploid species.** *Evolution* 2002, **56(6):**1101-1109.
18. Stouthamer R, Breeuwer JAJ, Hurst GDD: **Wolbachia pipientis: microbial manipulator of arthropod reproduction.** *Annual Review of Microbiology* 1999, **53:**71-102.
19. Schofield P: **Spatially explicit models of Turelli-Hoffmann Wolbachia invasive wave fronts.** *J Theor Biol* 2002, **215(1):**121-131.
20. Flor M, Hammerstein P, Telschow A: **Wolbachia-induced unidirectional cytoplasmic incompatibility and the stability of infection polymorphism in parapatric host populations.** *J Evol Biol* 2007, **20(2):**696-706.
21. Wade MJ, Stevens L: **The effect of population subdivision on the rate of spread of parasite-mediated cytoplasmic incompatibility.** *Journal of Theoretical Biology* 1994, **167:**81-87.
22. Guillaume F, Rougemont J: **Nemo: an evolutionary and population genetics programming framework.** *Bioinformatics* 2006, **22(20):**2556-2557.
23. Turelli M, Hoffmann AA: **Cytoplasmic incompatibility in Drosophila simulans: dynamics and parameter estimates from natural populations.** *Genetics* 1995, **140(4):**1319-1338.
24. Lehmann L, Bargum K, Reuter M: **An evolutionary analysis of the relationship between spite and altruism.** *J Evol Biol* 2006, **19:**1507-1516.
25. Lehmann L, Rousset F, Roze D, Keller L: **Strong reciprocity or strong ferocity? A population genetic view of the evolution of altruistic punishment.** *American Naturalist* 2007, **170:**21-36.
26. Whitlock MC: **Selection, load and inbreeding depression in a large metapopullation.** *Genetics* 2002, **160(3):**1191-1202.
27. Engelstädter J, Charlat S: **Outbreeding selects for spiteful cytoplasmic elements.** *Proceedings of the Royal Society B-Biological Sciences* 2006, **273(1589):**923-929.
28. Rasgon JL, Scott TW: **Impact of population age structure on Wolbachia transgene driver efficacy: ecologically complex factors and release of genetically modified mosquitoes.** *Insect Biochemistry and Molecular Biology* 2004, **34(7):**707-713.
29. Frankham R: **Effective population size/adult population size ratios in wildlife: a review.** *Genetical Research* 1995, **66(2):**95-107.
30. Dobson SL: **Reversing Wolbachia-based population replacement.** *Trends Parasitol* 2003, **19(3):**128-133.
31. Turelli M, Hoffmann AA: **Microbe-induced cytoplasmic incompatibility as a mechanism for introducing transgenes into arthropod populations.** *Insect Mol Biol* 1999, **8(2):**243-255.

Appendix

In this appendix, we provide a description of our model predicting the dynamics of CI-infections in structured populations. Infection is modelled as a two-state variable, with each individual being in one of the two states: "infected" or "uninfected". We designate by p the frequency of infected individuals among hosts in the population (hereafter referred to simply as the "infection frequency"). Because selection depends on mating frequencies and thus on the interaction between the sexes, we distinguish infection frequency in females, designated by p_{f}, from that in males, designated by p_{m}. We assume that infection is purely maternally transmitted and that organisms are diploid; since the sex-ratio is even we have $p_{\mathrm{f}} = p_{\mathrm{m}}$. Therefore, the change in frequency of infection is sufficiently described by the average change in frequency of infection in the population among female hosts, p_{f}, which can be written as

$$\Delta p_{\mathrm{f}} = \mathrm{E}\left[w_{ij}p_{\mathrm{f}(ij)}\right] - p_{\mathrm{f}}, \tag{1}$$

where w_{ij} is the expected number of successful offspring of female i reproducing in deme j, $p_{\mathrm{f}(ij)}$ is the frequency of infection in this female (0 or 1), and $\mathrm{E}\left[\cdot\right]$ denotes the average over all demes in the population and all females within a deme. Since the population is assumed to be of constant size, the mean fitness is equal to one ($\mathrm{E}\left[w_{ij}\right] = 1$). To keep our analysis simple, we assume that w_{ij} has a common form for all demes (i.e., the environment is homogeneous) and that the number of demes in the population is very large.

The fitness w_{ij} of female i in deme j depends on its expected number of offspring produced (i.e., fecundity), itself a function of infection frequency in female i and in males of deme j. In order to obtain an explicit expression for fecundity, we extend the expression of parasite fecundity for panmictic population given by Turelli (1994), and, without loss of generality, we consider only the fecundity of individuals relative to the fecundity of uninfected individuals. The relative fecundity of female i in deme j will be written as $1 + f_{ij}$, where f_{ij} represents the decrement of fecundity due to infection, composed of the effect of the genotype of the female herself and the genotype of the males with which she has mated. According to our life cycle assumptions (see main text), the relative fecundity is given by

$$1 + f_{ij} = 1 - Cp_{\mathrm{f}(ij)} - B(1 - p_{\mathrm{f}(ij)})p_{\mathrm{m}(j)}, \tag{2}$$

where $p_{\mathrm{m}(j)} = \frac{1}{N}\sum_i p_{\mathrm{m}(ij)}$ is the average frequency of infected males in deme j, and $p_{\mathrm{m}(kj)}$ is the frequency of infection (0 or 1) in male k from deme j. Averaging eq. 2 over all females

within deme j gives

$$
\begin{aligned}
1 + f_j &= \frac{1}{N} \sum_i (1 + f_{ij}) \\
&= 1 - C p_{\mathrm{f}(j)} - B(1 - p_{\mathrm{f}(j)}) p_{\mathrm{m}(j)},
\end{aligned}
\tag{3}
$$

where $p_{\mathrm{f}(j)} = \frac{1}{N} \sum_i p_{\mathrm{f}(ij)}$ is the average frequency of infected females in deme j. Finally, the average of eq. 3 over all demes in the population is

$$
\begin{aligned}
1 + f &= \frac{1}{n_{\mathrm{d}}} \sum_j (1 + f_j) \\
&= 1 - C p_{\mathrm{f}} - B p_{\mathrm{m}} + B \mathrm{E}\left[p_{\mathrm{m}(j)} p_{\mathrm{f}(j)} \right],
\end{aligned}
\tag{4}
$$

where the number n_{d} of demes in the population is assumed to be very large ($n_{\mathrm{d}} \to \infty$), and $\mathrm{E}\left[p_{\mathrm{m}(j)} p_{\mathrm{f}(j)} \right]$ is the probability of sampling an infected male and an infected female from the same deme. Due to our assumption of an infinite number of demes, an average fecundity over all demes, excluding a single focal deme, converges to the average fecundity $1 + f$ over all demes in the population. For notational simplicity, we will thus use in the following calculations $1 + f$ for any average over all demes while excluding a single focal deme.

We now have all the elements to evaluate the fitness function w_{ij}. A number $(1 - m)(1 + f_{ij})$ of the offspring of female i from group j remain in group j and compete against an average number $(1 - m)(1 + f_j)$ of offspring produced in that group and an average number $m(1 + f)$ of immigrant offspring produced in different groups. A number $m(1 + f_{ij})$ of the offspring of female i from group j disperse and enter in competition against an average number $(1 + f)$ of offspring. Collecting all terms, the fitness of female i from group j becomes

$$
w_{ij} = \frac{(1 - m)(1 + f_{ij})}{(1 - m)(1 + f_i) + m(1 + f)} + \frac{m(1 + f_{ij})}{1 + f},
\tag{5}
$$

which is of the same form as the fitness function of previous and closely related models (Roze and Rousset, 2003, 2004; Lehmann et al., 2007)

We now invoke weak selection, an assumptions that is often endorsed in social evolutionary theory or population genetics (e.g., Hamilton, 1964; Nagylaki, 1993; Taylor, 1996; Roze and Rousset, 2003, 2004; Kirkpatrick et al., 2002; Grafen, 2006), and implies that the phenotypic effects of infection, C and B, are of small order δ, where δ can be thought off as

the largest of the two effects of infection on fitness. Therefore, f_{ij}, f_j and f will also be of order δ and recalling that for x small we have $y/(1+x) \simeq y(1-x)$, eq. 5 becomes

$$w_{ij} = 1 + f_{ij} - (1-m)^2 f_j - (1 - (1-m)^2)f + O(\delta^2), \tag{6}$$

where $O(\delta^2)$ is a remainder of order δ^2. Substituting the fitness function into eq. 1 yields for weak selection

$$
\begin{aligned}
\Delta p_{\text{f}} &= \text{E}\left[\{f_{ij} - (1-m)^2 f_j - (1-(1-m)^2)f\}p_{\text{f}(ij)}\right] + O(\delta^2) \\
&= \text{E}\left[f_{ij}p_{\text{f}(ij)}\right] - (1-m)^2 \text{E}\left[f_j p_{\text{f}(ij)}\right] - (1-(1-m)^2)\text{E}\left[f p_{\text{f}(ij)}\right].
\end{aligned} \tag{7}
$$

Inserting eqs. 2-4 into this expression and noting that $p_{\text{f}(ij)}^2 = p_{\text{f}(ij)}$, the first expectation appearing in eq. 7 becomes

$$\text{E}\left[f_{ij}p_{\text{f}(ij)}\right] = -Cp_{\text{f}} \tag{8}$$

while the second expectation in eq. 7 is given by

$$\text{E}\left[f_j p_{\text{f}(ij)}\right] = -C\text{E}\left[p_{\text{f}(j)}p_{\text{f}(j)}\right] - B\{\text{E}\left[p_{\text{m}(j)}p_{\text{f}(j)}\right] - \text{E}\left[p_{\text{m}(j)}p_{\text{f}(j)}p_{\text{f}(j)}\right]\}, \tag{9}$$

and finally

$$\text{E}\left[f p_{\text{f}(ij)}\right] = p_{\text{f}}\left(-Cp_{\text{f}} - B\{p_{\text{m}} - \text{E}\left[p_{\text{m}(j)}p_{\text{f}(j)}\right]\}\right). \tag{10}$$

Substituting the last three terms into eq. 7, the change in infection frequency becomes

$$
\begin{aligned}
\Delta p_{\text{f}} = -C\,p_{\text{f}} + (1-m)^2 &\left(C\,\text{E}\left[p_{\text{f}(j)}p_{\text{f}(j)}\right] + B\{\text{E}\left[p_{\text{m}(j)}p_{\text{f}(j)}\right] - \text{E}\left[p_{\text{m}(j)}p_{\text{f}(j)}p_{\text{f}(j)}\right]\}\right) \\
&+ p_{\text{f}}\left(1 - (1-m)^2\right)\left(Cp_{\text{f}} + B\{p_{\text{m}} - \text{E}\left[p_{\text{m}(j)}p_{\text{f}(j)}\right]\}\right), \quad (11)
\end{aligned}
$$

where the expectations of pairs and triplets of infection frequencies are probabilities that pairs and triplets of individuals carry the infection. These probabilities will be evaluated by following the rationale usually employed for evaluating probabilities of identity between homologous genes in the infinite island model of dispersal, where two infection strains sampled from different demes are considered as genealogically independent, and thus bear no identity by descent (e.g., Perrin and Mazalov, 2000; Whitlock, 2002; Cherry and Wakeley, 2003; Roze and Rousset, 2003). For simplicity, we will also use the notation $p_{\text{f}} = p_{\text{m}} = p$.

We start by evaluating identities between pairs of infection statuses. The probability $\text{E}\left[p_{\text{f}(j)}p_{\text{f}(j)}\right]$ that two females randomly sampled with replacement from the same deme are

both infected comprises two events. First, the two females can have inherited their infection from an infected common ancestor. In the infinite island model this implies that, when looking backwards in time, the ancestral lineages of the two females' infection strains have always stayed in the same deme and coalesced in a local ancestor. If we call the probability of this happening $F_{\mathrm{ST}}^{\mathrm{R}}$, the probability of sharing an infection by descent is $F_{\mathrm{ST}}^{\mathrm{R}}p$, where the weight of the infection's frequency in the population, p, expresses the probability that the common ancestor carried the infection. Alternatively, the two females can have inherited their infections independently. This event occurs with the probability of non-coalescence between the two females' infection strains, $1 - F_{\mathrm{ST}}^{\mathrm{R}}$, weighted by the probability of two independent individuals sampled at random from the population being infected, p^2. Hence, the total probability that both females share the same infection statuses is

$$\mathrm{E}\left[p_{\mathrm{f}(j)}p_{\mathrm{f}(j)}\right] = F_{\mathrm{ST}}^{\mathrm{R}}p + (1 - F_{\mathrm{ST}}^{\mathrm{R}})p^2, \tag{12}$$

where

$$F_{\mathrm{ST}}^{\mathrm{R}} = \frac{1}{N} + \left(\frac{N-1}{N}\right)F_{\mathrm{ST}}. \tag{13}$$

In this expression, $1/N$ is the probability of sampling twice the same female and F_{ST} is the probability of coalescence of two distinct infection lineage that is given by

$$F_{\mathrm{ST}} = \sum_{t=1}^{\infty}(1-m)^{2t}\left(1 - \frac{1}{N}\right)^{t-1}\frac{1}{N}, \tag{14}$$

where $(1-m)^{2t}$ is the probability that the two infection lineages remained in the same deme for at least t generations and its factor is the probability that the lineages coalesced precisely at that time. In practice, it is more convenient to evaluate F_{ST} from recurrence equations (e.g., Hartl and Clark, 1997; Perrin and Mazalov, 2000; Gillespie, 2004; Roze and Rousset, 2003), which is given at equilibrium by

$$F_{\mathrm{ST}} = (1-m)^2\left(\frac{1}{N} + \frac{(N-1)}{N}F_{\mathrm{ST}}\right), \tag{15}$$

Following a reasoning along the same lines as those detailed above, we can calculate the probability $\mathrm{E}\left[p_{\mathrm{m}(j)}p_{\mathrm{f}(j)}\right]$ that a randomly sampled pair of one male and one female both bear the infection. This probability is given by

$$\mathrm{E}\left[p_{\mathrm{m}(j)}p_{\mathrm{f}(j)}\right] = F_{\mathrm{ST}}p + (1 - F_{\mathrm{ST}})p^2. \tag{16}$$

We now turn to the evaluation of the identity between triplets infection statuses. The probability $\mathrm{E}\left[p_{\mathrm{m}(j)}p_{\mathrm{f}(j)}p_{\mathrm{f}(j)}\right]$ that one male and two females sampled with replacement from

the same deme all bear the infection comprises three events. First, with probability $F_3^{\mathrm{R}}p$ the three individuals share their infection statuses by common descent because all three gene lineages coalesce in a local common ancestor (with a probability called F_3^{R}) who bore the infection strain (with probability p). Second, the three individuals can all be infected because two of them share their infection by common local descent while the third individual is part of an immigrant lineage of the infected type. This occurs with probability $(F_{\mathrm{ST}}^{\mathrm{R}}+2F_{\mathrm{ST}}-3F_3^{\mathrm{R}})p^2$ where $F_{\mathrm{ST}}^{\mathrm{R}}$ is the probability of coalescence of genes in two females sampled with replacement in a deme and F_{ST} is the probability of coalescence of two gene, one in a male and one in a female. The probability F_3^{R} has to be subtracted in order to exclude cases in which not only one pair of strain lineages but all three coalesce. Third, with a probability complementary to the two just described, the three strains lineages do not coalesce, in which case they all bear the infection status with probability p^3. Thus,

$$\mathrm{E}\left[p_{\mathrm{m}(j)}p_{\mathrm{f}(j)}p_{\mathrm{f}(j)}\right] = F_3^{\mathrm{R}}p + \left(F_{\mathrm{ST}}^{\mathrm{R}} + 2F_{\mathrm{ST}} - 3F_3^{\mathrm{R}}\right)p^2$$
$$+ \left(1 - F_{\mathrm{ST}}^{\mathrm{R}} - 2F_{\mathrm{ST}} + 2F_3^{\mathrm{R}}\right)p^3, \quad (17)$$

where

$$F_3^{\mathrm{R}} = \frac{1}{N}F_{\mathrm{ST}} + \left(\frac{N-1}{N}\right)F_3, \tag{18}$$

is the probability of identity by descent between three strains, two of which are sampled with replacement and

$$F_3 = (1-m)^3 \left(\frac{1}{N^2} + \frac{3(N-1)}{N^2}F_{\mathrm{ST}} + \frac{(N-2)(N-1)}{N^2}F_3\right). \tag{19}$$

is the equilibrium probability of identity between three strains sampled without replacement from the same deme after dispersal.

Inserting all the probabilities of identity in state (eq. 12, eq. 16 and eq. 17) into the equation of infection frequency change (eq. 11), we obtain after simplification that

$$\Delta p = p(1-p)\left[-C(1-F_{\mathrm{ST}}) + B(1-m)^2\left(F_{\mathrm{ST}} - F_3^{\mathrm{R}}\right) + pB\{1 - g(N,m)\}\right]. \tag{20}$$

where

$$g(N,m) = 2F_{\mathrm{ST}} + (1-m)^2(F_{\mathrm{ST}} - 2F_3^{\mathrm{R}}). \tag{21}$$

From these equations we can express the condition for infection to spread when rare ($p \to 0$) in terms of the cost-to-benefit ratio as

$$\frac{C}{B} < (1-m)^2 \left(\frac{F_{\mathrm{ST}} - F_3^{\mathrm{R}}}{1 - F_{\mathrm{ST}}}\right). \tag{22}$$

References

Cherry, J. L. and J. Wakeley. 2003. A diffusion approximation for selection and drift in a subdivided population. *Genetics* 163:421–428.

Gillespie, J. H. 2004. Population genetics: a concise guide. Johns Hopkins, Baltimore & London.

Grafen, A. 2006. Optimization of inclusive fitness. *Journal of Theoretical Biology* 238:541–563.

Hamilton, W. D. 1964. The genetical evolution of social behaviour, I. *Journal of Theoretical Biology* 7:1–16.

Hartl, D. and A. G. Clark. 1997. Principles of population genetics. Sinauer, Massachusetts.

Kirkpatrick, M., T. Johnson, and N. Barton. 2002. General models of multilocus evolution. *Genetics* 161:1727–1750.

Lehmann, L., F. Rousset, D. Roze, and L. Keller. 2007. Strong reciprocity or strong ferocity? A population genetic view of the evolution of altruistic punishment. *The American Naturalist* 170:21–36.

Nagylaki, T. 1993. The evolution of multilocus systems under weak selection. *Genetics* 134:627–647.

Perrin, N. and V. Mazalov. 2000. Local competition, inbreeding and the evolution of sex-biased dispersal. *The American Naturalist* 155:116–127.

Roze, D. and F. Rousset. 2003. Selection and drift in subdivided populations: A straightforward method for deriving diffusion approximations and applications involving dominance, selfing and local extinctions. *Genetics* 165:2153–2166.

Roze, D. and F. Rousset. 2004. The robustness of Hamilton's rule with inbreeding and dominance: Kin selection and fixation probabilities under partial sib mating. *American Naturalist* 164:214–231.

Taylor, P. D. 1996. Inclusive fitness arguments in genetic models of behaviour. *Journal of Mathematical Biology* 34:654–674.

Turelli, M. 1994. Evolution of incompatibility-inducing microbes and their hosts. *Evolution* 48:1500–1513.

Whitlock, M. C. 2002. Selection, load and inbreeding depression in a large metapopulation. *Genetics* 160:1191–1202.