Genotype and sex-based host variation in behavior and susceptibility drives population disease dynamics

Genotype and sex-based host variation in behavior and susceptibility drives population disease dynamics

Lauren A. White[1,2], Jonathon A. Siva-Jothy[3], Meggan E. Craft[2], Pedro F. Vale[3]

1. National Socio-Environmental Synthesis Center SESYNC, 1 Park Place, Suite 300, Annapolis, MD 21401, USA.

2. Department of Veterinary Population Medicine, University of Minnesota, St Paul, MN 55126

3.Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Ashworth Labs, Charlotte Auerbach Road, EH9 3JT Edinburgh, UK.

Abstract

Host heterogeneity in pathogen transmission is widespread and presents a major hurdle to predicting and minimizing disease outbreaks. Using *Drosophila melanogaster* infected with *Drosophila* C virus as a model system, we integrated experimental measurements of social aggregation, virus shedding, and disease-induced mortality from different genetic lines and sexes into a disease modelling framework. The experimentally measured host heterogeneity produced substantial differences in simulated disease outbreaks, providing evidence for genetic and sex-specific effects on disease dynamics at a population level. While this was true for homogeneous populations of single sex/genetic line, the genetic background or sex of the index case did not alter outbreak dynamics in simulated, heterogeneous populations. Finally, to explore the relative effects of social aggregation, viral shedding and mortality, we compared simulations where we allowed these traits to vary, as measured experimentally, to simulations where we constrained variation in these traits to the population mean. In this context, variation in infectiousness, followed by social aggregation, was the most influential component of transmission. Overall, we show that host heterogeneity in three host traits dramatically affects population-level transmission, but the relative impact of this variation depends on both the susceptible population diversity and the distribution of population-level variation.

Key words: *Drosophila melanogaster*, disease transmission, social aggregation, virus shedding, contact networks, disease modelling, transmission heterogeneity

Introduction

Individual heterogeneity in host traits affecting disease transmission has major consequences for the predictability and severity of outbreaks of infectious disease, and in extreme cases can lead to 'superspreaders' or 'supershedders' of infection [1–3]. An individual's transmission potential can be described as a function of: (1) its rate of contact with susceptible individuals, (2) the likelihood of that contact resulting in infection, and (3) the length of time that individual remains infectious [4,5]. It is therefore important to understand how common sources of variation, such as host genetic background and sex, may contribute to the variance in these traits and how individual variation may scale up to population level disease dynamics [4,6,7].

Disease dynamics may be disproportionately driven by individuals with extreme behavioural and physiological traits including social aggregation, pathogen shedding, or in the host's SĺAability to resist or tolerate the infection. For example, sex differences in immunity [8] or nutritional and thermal effects on host behaviours [9,10] can lead to differences in hosts' ability to tolerate infection and consequently increase transmission rates. Similarly, there are also examples of genetic differences driving the extent of pathogen shedding [11] and behaviours that mediate contact between infected and susceptible individuals [12,13]. Quantifying these relevant behavioural, physiological and immune traits and their interactions remains tremendously challenging, particularly in wild or natural disease settings [4].

One potentially useful approach is experimentally infecting model systems under controlled laboratory settings in order to quantify the roles of physiological and behavioural host heterogeneity on pathogen transmission [12,14,15]. This experimental approach offers the advantage of providing an experimentally tractable framework to partition the variance in individual transmission among a range of behavioural, physiological and immune phenotypes [4], while minimising environmental variation and allowing highly replicated measurements of individual host traits. However, such studies may be limited in their ability to extrapolate the effects of measured heterogeneity at the level

of individual hosts to population-level epidemic dynamics. Mathematical modelling is a useful tool to efficiently test different hypotheses and infer patterns across scales [16], but many theoretical studies often rely on assumptions about the level of heterogeneity in host traits, in the absence of empirical data [4,5]. A helpful approach is therefore to use mathematical modelling of epidemiological dynamics whereas many parameters as possible are informed by experimental data measured on individual hosts in controlled laboratory settings.

Here we combine experimental data and a simulation approach to test how population-level disease transmission dynamics are affected by experimentally measured levels of variation in pathogen shedding, lifespan following infection and social aggregation. We previously measured individual-level variation in behavioural and physiological traits that are relevant to pathogen transmission in the fruit fly (*Drosophila melanogaster)* when infected with its viral pathogen *Drosophila* C Virus (DCV) [13,17]. DCV is a horizontally transmitted ssRNA virus of Drosophila. While relatively little is known about DCV dynamics in the wild, it appears to be common as a low-level persistent infection with apparently little pathology among several species of *Drosophila* [18,19]. Following what is presumably a predominantly fecal-oral route of transmission, DCV replicates in the fly's reproductive and digestive tissues leading to intestinal obstruction, lower metabolic rate and reduced locomotor activity [20–22]. Some experimental work has also shown that cannibalism of infectious fly cadavers is a viable route of transmission, but it is unknown how common this transmission route is in the wild [23]. Previously, we observed sex-based and genetic-based variation in both locomotor activity and social aggregation following DCV infection [13]. We also showed that fly genetic background, sex and female mating status significantly influenced infected lifespan, viral growth, virus shedding, and viral load at death [17]. These experiments leveraged genetic and sex-specific sources of variation in three traits that likely affect individual transmission potential of DCV: the degree of group-level social aggregation (as an indicator of potential contact rate); how much DCV each individual sheds into its environment (as a proxy measure of infectiousness); and mortality rate (which defines the duration of infection).

92 In the present study, we explore the interactions of social aggregation, viral shedding, and mortality
93 on pathogen transmission when we: (1) vary population means of these traits; (2) vary the individual
94 traits of the index case; and (3) constrain the variance of these traits in the population at large. First,
95 we asked if genetic and sex-specific variation in the population means of social aggregation, virus
96 shedding, and duration of infection – as measured in a lab setting – would result in different predicted
97 epidemics in theoretical populations. By comparing simulated epidemics in host populations
98 comprised of a single sex and one genetic background, we isolated genetic and sex-specific sources of
99 variation in disease transmission. Second, to test the relative importance of the index case vs. group
100 composition, we simulated epidemics in populations where the index case's traits were sampled from
101 a larger phenotypic distribution, including males and females from all ten genetic backgrounds. Third,
102 to test the relative importance of variation in specific host traits on epidemic dynamics, we compared
103 epidemic dynamics of populations exhibiting experimentally-measured levels of variation in social
104 aggregation, viral shedding and mortality, to populations where we constrained variation in these
105 traits to the population mean.
106

Methods

Simulation model

We developed an individual-based, stochastic, discreate time model that tests how experimentally measured variation in host social aggregation, mortality, and viral shedding in *D. melanogaster* translates to differences in disease dynamics. The simulated contact networks underlying this model were generated from degree distributions derived from experimental measurements of social aggregation specific to the sex (σ) and genetic line (g) present in the simulated population. Using a susceptible-infected-removed (*SIR*) process, we simulated direct transmission of DCV in a closed population with no births and where only infected individuals die [24]. We did not include background (i.e. non-disease related) mortality. Note that these transmission processes are consistent with other agent-based models that encompass contact heterogeneity[25].

Let the time step be equal to one day, $S(t)$ equal the number of susceptible hosts at time *t*, and $I(t)$ equal the number of infectious hosts at time *t*. The total number of hosts, $N(t)$, in the population at time *t* is represented by: $N(t) = S(t) + I(t)$. The number of susceptible (*S*) and infected (*I*) individuals at the next time step is given by:

$$S(t+1) = S(t) - \sum_{i=1}^{S(t)} \sum_{j=1}^{I(t)} \beta_{ij}(t) s_i(t) i_j(t)$$

$$I(t+1) = I(t) + \sum_{i=1}^{S(t)} \sum_{j=1}^{I(t)} \beta_{ij}(t) s_i(t) i_j(t) - \alpha_j(\sigma, g) i_j(t)$$

Here $s_i(t)$ is a vector of susceptible individuals at time *t*, and $i_j(t)$ is a vector of infected individuals at time *t*. Therefore, the summations in both equations above iterate over individuals and not time steps.

The processes of mortality $\alpha_j(\sigma, g)$ and transmission (β_{ij}) were individual-specific reflecting the

covariates of sex (σ) genetic line (g). More specifically, the transmission between a susceptible individual (s_i) and infectious host (j) is given by:

$$\beta_{ij}(t) = \kappa_j(\sigma, g)\eta\tau x_{ij}(t)$$

where $\kappa_j(\sigma, g)$ represents the infectiousness of infectious host j and x_{ij} represents whether or not an edge exists in the network between individuals s_i and j $\left(x_{ij}(t) = \begin{Bmatrix}1\\0\end{Bmatrix}\right)$. Because of the uncertainty surrounding the DCV transmission process, we also include scaled infectiousness (η) and transmission efficiency of the pathogen (τ) as components of $\beta_{ij}(t)$, which we discuss in further detail below.

For each susceptible individual s_i at each time step, transmission was a stochastic process governed by a Bernoulli draw based on the value of $\beta_{ij}(t)$. Likewise, for each infectious individual j, mortality was a stochastic process stochastic processes based on a Bernoulli draw for the value of $\alpha_j(\sigma, g)$. Individuals removed during the mortality process no longer contributed to transmission dynamics.

Experimental data distributions: Measuring social aggregation, viral shedding and mortality rate in infected *D. melanogaster*

We used experimental measurements of host social aggregation, mortality, and viral shedding from *D. melanogaster* infected with DCV (Figure 1a-c; note the heterogeneity among genotypes) to test how the sex-specific and genetic variation translates to differences in disease dynamics. An in-depth analysis of these experimental data has been carried out previously, showing substantial genotype-by-sex interactive effects on each of these traits [13,17]. Briefly, we established systemic infections with DCV in males and females of ten lines (Table 1) from the *Drosophila* Genetic Resource Panel (DGRP) [26] and measured a number of traits including social aggregation [20], the infected lifespan and the viral shedding of each line-by-sex combination [17].

Here, we focus on the frequency distributions of these data for each fly line and sex (Figure 1), as the simulations described below were parameterized using these experimentally derived

156 distributions. Of particular note is the distribution of viral shedding (Figure 1b), which showed
157 substantial zero-inflation, due to many flies not shedding DCV in detectable quantities despite being
158 infected.
159
160 **Social aggregation and contact network degree distribution**
161 Social aggregation was measured by calculating the nearest neighbour distance (NND) from a
162 photograph of groups of ten to twelve flies of the same genetic background, sex and infection status,
163 in 55mm Petri dishes [13]. In accordance with other studies of *D. melanogaster* social aggregation
164 [45], photos were taken of fly groups in Petri dishes following 30 minutes of acclimation to ensure
165 minimal fly activity. Social aggregation was measured in n = 14–16 replicate groups of 12 flies for
166 every combination of genetic background and sex (580 groups of flies in total).
167 The dynamics of faecal-oral DCV transmission are poorly understood [27,28], but the virus
168 readily proliferates through laboratory stocks of *Drosophila* [29]. To account for this uncertainty in
169 transmission mode and to assess the relative importance of possible direct transmission routes, we
170 considered three threshold radii (10, 15 or 20mm) for feasible transmission. For each of these
171 thresholds, the qualifying neighbours for each focal individual was calculated using the coordinates of
172 each fly generated with the ImageJ multipoint tool.
173 To generate a simulated contact network reflecting contact rates of different phenotypes, we
174 started by creating empirical contact networks where an individual (node) shared an edge in the
175 network if they appeared within the prescribed threshold radius of the focal fly. Importantly, using
176 social aggregation as a proximate measure of contact rate assumes the likelihood of contact with DCV
177 is proportional to an individual's proximity to an infected fly. Using the number of neighbours within
178 this radius for each fly (i.e., unweighted degree centrality), we derived an empirical degree
179 distribution for each genetic line and sex combination. From this empirical degree distribution, we
180 sampled 1000 times with replacement to generate a larger degree distribution representing more
181 individuals. To produce a random graph with this given degree sequence, we then used the
182 samp_degseq function from the igraph package [30]. Note that we resampled if the degree sequence

summed to be odd. This produced a network where the mean degree (rather than network density) was maintained between experimental and simulated populations.

Infectiousness. We estimated infectiousness (κ_j) for any given infected individual, j, from our experimental measurements of viral shedding [17]. Viral shedding was measured by housing single infected flies in 1.5ml Eppendorf tubes for 24 hours, removing the fly, washing out the tube with 50µl of TRI-reagent to preserve viral RNA, and freezing this sample at -70°C to await RT-PCR and qPCR. Each combination of sex and genetic background consisted of a minimum of 20 replicate flies, with most combinations consisting of 32-38 shedding samples [17].

The untransformed distribution of this data was highly skewed and zero-inflated, with some rare flies shedding exceedingly high viral titres (i.e. supershedders)—over two orders of magnitude greater than the population mean—and others not shedding any virus at all (within the technical limit of detection). To account for this disparity, we used the natural log to transform our viral load shed distribution and then normalized values by the greatest amount of virus shed. This transformation yielded a distribution constrained between 0 and 1 with a median value of 0 and a mean value of 0.23. With this transformed distribution, only extreme supershedders at the upper end of the distribution would ensure a high probability transmission, with all other individuals had a probability much less than one.

Since the amount of virus needed to ensure DCV transmission is unclear, we also considered a 'scaled infectiousness' (η) parameter to explore what would happen if average or non-zero shedders could also shed enough to ensure infection. This scenario was implemented by multiplying our measure of infectiousness (κ_j) by 2. This step expanded the range of the transformed experimental distribution from 0 to 2. Note that for the Bernoulli trial determining whether a transmission event had occurred, the final transmission probability () was than capped at a maximum value of 1.

Finally, because the dosage and viability of DCV in the environment remain unclear, we included a transmission efficiency (τ) parameter in our model to account for this uncertainty. The three levels, $\tau = 0.1$, 0.5, or 1, altered infectiousness and correspond to 10, 50, and 100% probability

of transmission given contact. Both scaled infectiousness (η) and transmission efficiency (τ) were held constant in simulations unless specifically mentioned.

Mortality rate. DCV results in death for infected flies, making our experimental measurement of the time between inoculation and death an ideal measure of mortality rate. Infected lifespan was measured by housing single flies in standard Lewis medium vials following systemic DCV infection and monitored daily until death. For eighteen of twenty sex and genetic background combinations, the lifespan following infection was measured for n=17-20, two combinations consisted of n=13 and n=15 flies [17]. For simulations, we calculated mortality rate, $\alpha_j(\sigma, g)$, as the inverse of experimentally-measured disease-related mortality for a given sex and genetic background.

Simulation factorial design

The effects of all parameters on outbreak dynamics were tested in a full-factorial design. For each parameter set, 500 simulations were conducted for a population of 1000 individuals over the course of 1000 time steps (Tables 1-3). A wide variety of outbreaks of infectious disease were produced by different combinations of these parameters. To avoid datasets becoming predominated by fadeout, we have presented the outbreaks in populations defined by a set of parameters (r=15mm, τ =1, η=2) most conducive to outbreaks of infectious disease. Key metrics to measure outbreak dynamics included: fadeout probability, maximum number of infected individuals, outbreak duration, and time to maximum number of infected individuals. Fadeout probability represents the probability of an outbreak stochastically dying out [31]; in this case, we define it as the proportion of simulations where DCV fails to spread beyond the index case. We use R_0 as a measure of the number of secondary cases of infection caused by the index case for the duration of the simulation. Code to conduct these simulations was written in *R* (Version 3.4.4) and is available at:
https://github.com/whit1951/Drosophila

Random forest analysis

Parsing out the effects of individual variables in simulation modelling can be challenging because of collinear effects and sensitivity of frequentist measures of significance to sample size. To further a descriptive discussion of our simulation results, we have used random forest analysis – a machine learning method that can handle complex, non-linear relationships between model inputs and outputs, as well as potential collinearity between covariates [32]. Random forest analysis is a recursive partitioning method that combines the predictions from numerous fittings of classification or regression trees to the same set of data [32,33]. A higher mean decrease in accuracy correlates with higher variable importance, i.e., more predictive power is lost if this variable is excluded from the analysis. For all three simulation experiments, we analysed outputs of: fadeout probability (whether the infection spread beyond initially infected individual), maximum prevalence, outbreak duration, and R_0 (the number of secondary cases resulting from a single infectious individual in an entirely susceptible population). A detailed description of the analyses can be found in supplementary information and Figures S1-S3.

Results

Simulation results

Overall, our findings were robust to changes in various parameter combinations (Table 1). Threshold radius had strong effects on maximum prevalence but was not as strong predictor of a predictor of outbreak likelihood (Figures S1-S4). Here, we present results for a threshold radius of 15mm, a transmission efficiency of 1, and a scaled infectiousness of 2, which were generally representative of most parameter spaces. Summary figures for every parameter combination are presented in Figures S4-S15.

Theoretical simulation #1: We scaled-up the experimental degree distributions for males and females of our ten genetic backgrounds to a theoretical population size of 1000. In each simulated population, flies were of the same sex and genetic background. We allowed infectiousness, duration of infection, and social aggregation to vary based on experimental measurements for each combination of sex and genetic background (Table 1). For each individual simulation, we generated a new network from the scaled-up degree distribution, and randomly selected an individual from the network to start as the index case.

Individual variation in host infectiousness, social aggregation, and mortality rate produced variation in population-level, pathogen transmission dynamics.

The variation in experimental treatment groups produced distinct outbreaks of infectious disease in populations comprised solely of one genetic background and sex (Figure 2a-d). This finding held true when comparing both genetic lines and sexes. For example, the median outbreak size for line 373 females was ~200 flies compared to ~1 fly for line 373 males. In contrast the median outbreak size for line 818 females was ~1 fly, but approached ~500 flies for line 818 males. Random forest analysis suggested that the two top predictors for outbreak likelihood were genetic and sex-specific variation (Figure S1). Given a successful outbreak, host genetic and sex-specific variation also affected the

maximum number of infected individuals at any given time step (Figures 2c & S1) and outbreak duration (Figures 2d & S1). However, host genetic background and sex were less important than the threshold radius used to derive social network degree distribution for both outcomes (Figure S1) and less important than transmission efficiency for predicting the maximum number of infected individuals (Figure S1).

Theoretical simulation #2: Many natural host populations have highly variable levels of genetic diversity which can significantly affect host-pathogen dynamics [34]. To test the relative importance of trait differences among potential index cases, we simulated populations where males and females of all ten genetic backgrounds were combined in equal proportion. More specifically, the simulated, scaled-up populations of 1000 individuals were comprised of 20 sub-populations each containing 50 sampled individuals drawn from the larger experimental distribution for each respective line/sex combination. Individuals maintained their respective experimentally measured distributions for aggregation, infectiousness, and duration of infection according to their genetic background and sex combination. A connected network of these sub-populations was created by sampling an expected degree for each node based on its subpopulation traits and then using the samp_degseq function from the igraph package to create a random graph with the given degree sequence as described in the Methods [30]. Thus, flies with different covariate traits (as simulated from sampling from their respective experimental data distributions) could be connected in the network. These simulated populations therefore reflect a relatively diverse population. We then varied which genetic background and sex combination served as the index case (Table 1). We conducted 500 replicates per index case type. For each recorded replicate, the traits of the simulated population were resampled, and a new network was generated.

Effects of the index case were outweighed by heterogeneity in the susceptible population.

The genetic background or sex of the index case did not alter outbreak dynamics in diverse populations where 20 experimental treatment groups (all genotype by sex combinations) were equally sampled to create a heterogeneous population (Figure 3 & S2). This was true for all outbreak

parameters (Figure 3 & S2). Based on the random forest analysis, threshold radius and transmission efficiency were the top two predictors for fadeout probability, maximum number of infected individuals, outbreak duration, and R_0 (Figure S2).

Theoretical simulation #3: To determine the relative importance of experimentally observed variation in social aggregation, viral shedding, and disease-related mortality on disease transmission in a heterogeneous population, we simulated heterogeneous populations derived from the variation seen across all genetic backgrounds and both sexes. To determine the effect of population-level variation, we iteratively constrained the variation in each three host traits to the population's mean. During these simulations, the unconstrained traits were free to vary according to their experimentally determined distributions (Table 1). For example, to understand at the effect of variation in social aggregation in isolation, we constrained social aggregation to take on the experimentally determined mean degree distribution of the entire heterogeneous population but allowed viral shedding and mortality rate to vary according to their experimentally-measured distributions across all genetic backgrounds and both sexes. In the case of degree of the network, we rounded this value to ensure a whole number, which is essential for contact network formation (e.g., an individual cannot have 2.5 contacts). We also considered interactions between variability of these three traits (Table 1).

Variation in infectiousness increased fadeout probability and decreased maximum prevalence of successful outbreaks, but increased outbreak duration.

Constraining the infectiousness of a population to the mean (0.23, 0.46 for scaled infectiousness (η) levels 1 and 2, respectively) of the experimentally measured distribution increased the outbreak severity (Figure 4a), made outbreaks 2-fold more likely (Figure 4b), more than doubled the maximum prevalence (Figure 4a,c), and persisted in the population for longer (Figure 4a,d). Limiting variation in infectiousness also made outbreaks more predictable, reducing the variance of the time taken to reach the maximum number of infected individuals (Figure 4d). According to the random forest

analysis, variation in infectiousness was the top predictor for whether or not an outbreak spread beyond the initially infected individual (Figure S3).

Variation in social aggregation did not influence fadeout probability but made outbreaks more severe

When social network degree distribution of simulated populations was confined to the mean of the experimental data (2, 3 and 4 for threshold radii of 10, 15 and 20mm respectively), outbreaks became less severe (Figure 3a) compared to simulations based on the complete degree distribution. Simulated DCV spread to fewer individuals (Figure 4c) and was quicker to die-out than in simulations where infectiousness, social aggregation, and mortality varied freely (Figure 4d).

Variation in disease-related mortality did not affect epidemic outcomes.

When constrained to the mean of the experimental data (13.6 days), we found disease-related mortality had little to no effect on any aspect of disease outbreak (Figure 4). This is supported by the random forest analysis which identified variation in mortality rate as the least important predictor across outbreak metrics (Figure S3).

Variation in infectiousness, followed by social aggregation, was the most influential component of transmission.

An increase in the maximum number of infected individuals was only seen when variation in infectiousness was constrained. Interestingly the same effect was seen in simulations where other traits are constrained alongside virus shedding, despite this differing substantially from the effects of social aggregation and mortality rate when constrained alone (Figures 4 & S3). A similar, overruling effect was seen when social aggregation and mortality rate were constrained simultaneously, and virus shedding varied freely; outbreak dynamics were similar to the cases where only aggregation is constrained (Figures 4 & S3).

Discussion

Here, we investigated how host genetic background and sex may contribute to the variance in social aggregation, infectiousness and mortality and how this variation may scale up to population level disease dynamics. We found substantial between-individual differences in pathogen transmission, constituting genetic and sex-specific variation in transmission potential. Crucially, in relatively homogenous populations comprised of single sex and genotype combinations, heterogeneity in the index case produced major differences in population-level outbreak dynamics, including making outbreaks more likely, broader reaching, and longer lasting. However, variation in the index case's transmission potential exerted little influence over population-level outbreak dynamics in diverse host populations. We also found that population-level variation in social aggregation, virus shedding, and disease-related mortality affected outbreak dynamics in starkly contrasting ways. This effect appeared to be linked to the population-level distribution of each respective host trait, with factors such as skewness and zero-inflation influencing how variation in each trait affected outbreak dynamics.

In simulation experiment #1, males from the RAL-818 genetic background were not only more likely to start an outbreak of infectious disease, but these outbreaks were also more severe than in other populations. This suggests these males represent a class of individuals with a high transmission risk. Interestingly, high-risk males are seen in a number of host-pathogen systems [35,36]. While high-risk male classes can be produced by a range of traits pertaining to sex-specific ecology or physiology, their occurrence across systems is likely driven by sexual selection shaping male traits affecting transmission [37]. For example, in the yellow-necked mouse, *Apodemus flavicollis*, males are thought to be a high-risk class due to a range of sex differences in their immune response, home range and contact rates [35]. Moreover, as male *Drosophila* exhibit a number of other traits with the potential to alter their transmission potential, such as male-male fighting [38], the transmission risk of RAL-818 males could increase further. Focussing on classes of high-risk individuals is a more pragmatic approach to reducing the effect of heterogeneity in transmission potential, requiring less intensive monitoring protocols [4]. Additionally, as classes of individuals are identified using ranges of physiological or behavioural traits, classes are potentially more generalisable to other host-pathogen

systems (e.g. sex, social dominance). Many studies of transmission heterogeneity in natural systems focus on using either behavioural or physiological traits to infer transmission dynamics and identify high-risk individuals [2,4]. Our results highlight the importance of disentangling the relative contributions made by behavioural and physiological traits together in order to infer variation in transmission potential.

High-risk individuals, such as superspreaders, present a challenge to current methods of disease control because they are capable of starting outbreaks of infectious disease that are difficult to predict and amplifying them once transmission begins [39,40]. Pre-emptively identifying high-risk individuals is therefore a major aim of epidemiology and disease ecology. However, in the second theoretical experiment we conducted, we found that starting outbreaks with individuals that differed in transmission potential did not affect outbreak dynamics when susceptible populations are genetically diverse. Our results therefore suggest outbreaks are not solely driven by the traits of rare, high-risk individuals, but are also affected by the traits of the susceptible population. High-risk individuals were unable to cause explosive outbreaks of infectious disease when surrounded by low-risk individuals as presumably, once infected, low-risk individuals failed to transmit disease to the rest of the population. Similar transmission dynamics have also been observed in laboratory populations of the social spider, *Stegodyphus dumicola*, where transmission of a bacterial pathogen was affected by the boldness of the index case and the individuals it interacted with [14], but ultimately traits of the index case did not alter transmission dynamics compared to the collective traits of the susceptible population. Together with our results, these findings do not suggest diversity in the susceptible population is a universal buffer to the effects of between-individual heterogeneity in disease transmission. Instead, this work highlights the necessity to characterise population diversity in the context of social interactions and networks as these may determine the relevance of this diversity. Population-level diversity is particularly important in host-pathogen systems where behavioural changes occur following infection. In populations of the guppy, *Poecilia reticulata*, for example, male, but not female, sociality has been shown to increase following infection. As a result, females social males are more likely to interact with, and infect, females [12]. There are many traits across

species that bias social interactions, such as sexual receptivity or personality type [41]. Should these traits bias contact between transmission classes, this may explain why social and transmission networks rarely match.

Extreme phenotypes often play a key role in between-individual heterogeneity in disease transmission. However, being a relative term, 'extreme' phenotypes are defined by population-level variation. Constraining population-level variation in the amount of virus shed following infection to the population mean increased outbreak likelihood and severity. This was likely a result of the huge zero-inflation of the distribution of virus shedding, where many infected individuals did not shed virus. These individuals, previously termed 'supersponges' [42], represent the left-most extreme of the population distribution, and bore no transmission risk. While some of the individuals that do not transmit infection may simply not get any transmission opportunity, others may be supersponges and therefore incapable of transmitting disease. The presence of supersponges also demonstrates the importance of measuring variation in both behavioural and physiological traits when seeking to understand heterogeneity in disease transmission. Characterising extreme forms of population-level variation, particularly in natural systems where experiments are less controlled, should certainly be prioritised in order to understand individual heterogeneity in disease transmission.

An important caveat of our results is that because we did not measure social aggregation, virus shedding and lifespan simultaneously we cannot account for how they might covary within individuals. We therefore allow them to co-occur in hosts randomly, which may not reflect associations produced in nature or potential combinations of traits that are not likely due to physiological or evolutionary constraints. This is particularly true for how we estimated contact behaviour from social aggregation arenas containing 10-12 flies and measuring 55mm wide. For our simulations, we scaled-up these smaller populations to create theoretical populations of 1000 individuals. This approach was required by the experimental demands of measuring social aggregation, although it is known that social aggregation changes may change with population size and sex ratio [43,44].

Threshold radius was a singularly important parameter across our theoretical experiment. Understanding how distance affects pathogen transmission or definitions of what constitutes a contact remains poorly described in many host-pathogen systems [7]. Moreover, real networks may have different structures not accounted for here, such as a modular structure which has been shown to facilitate or prevent the spread of disease [44,45]. As our social aggregation data comes from Petri dishes containing only males or females from a single genetic background, we cannot account for how aggregation might change in more diverse and larger populations[43].

Our work bears a number of consequences for understanding how between-individual heterogeneity in disease transmission is determined and how it could affect outbreak dynamics. We show that variation in key individual traits can dramatically affect population-level transmission, surmounting to genetic and sex-specific variation in transmission potential. Importantly, the influence of this variation is dramatically affected by susceptible population diversity and the distribution of population-level variation. These results support the observations of other systems that suggest the traits of susceptible individuals can exert significant influence over transmission. This is particularly relevant to populations with low genetic diversity, such as agricultural monocultures, as this lack of diversity increases the risk of explosive outbreaks [46]. Our work posits the merits of integrating data collected in highly controlled laboratory experiments with simulations capable of extrapolating this information to larger populations.

Funding

J.A.S-J was funded by a NERC E3 DTP PhD studentship awarded to the University of Edinburgh. L.A.W. was funded by the National Science Foundation (GRFP-00039202 and DEB-1701069), the University of Minnesota Informatics Institute, and the National Socio-Environmental Synthesis Center (SESYNC) under funding received from the National Science Foundation DBI-1639145. J.A.S-J and L.A.W. received funding for a research exchange from the Infectious Disease Evolution Across Scales RCN funded by NSF. M.E.C. was funded by National Science Foundation (DEB-

1413925, 1654609, and 2030509). The authors acknowledge the Minnesota Supercomputing Institute (MSI) at the University of Minnesota for providing resources that contributed to the research results reported within this paper. URL: http://www.msi.umn.edu. P.F.V was supported by a Branco Weiss fellowship (https://brancoweissfellowship.org/) and a Chancellor's Fellowship (School of Biological Sciences, University of Edinburgh).

References

1. Craft ME. 2015 Infectious disease transmission and contact networks in wildlife and livestock. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **370**, 20140107-. (doi:10.1098/rstb.2014.0107)

2. White LA, Forester JD, Craft ME. 2017 Using contact networks to explore mechanisms of parasite transmission in wildlife. *Biol. Rev.* **92**, 389–409. (doi:10.1111/brv.12236)

3. Lloyd-Smith JO, Schreiber S, Kopp P, Getz W. 2005 Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359. (doi:10.1038/nature04153)

4. VanderWaal KL, Ezenwa VO. 2016 Heterogeneity in pathogen transmission: mechanisms and methodology. *Funct. Ecol.* **30**, 1606–1622. (doi:10.1111/1365-2435.12645)

5. McCallum H *et al.* 2017 Breaking beta: deconstructing the parasite transmission function. *Philos. Trans. B* **372**, 20160084. (doi:10.1098/rstb.2016.0084)

6. Hawley DM, Altizer SM. 2011 Disease ecology meets ecological immunology: understanding the links between organismal immunity and infection dynamics in natural populations. *Funct. Ecol.* **25**, 48–60. (doi:10.1111/j.1365-2435.2010.01753.x)

7. White LA, Forester JD, Craft ME. 2018 Covariation between the physiological and behavioral components of pathogen transmission: host heterogeneity determines epidemic outcomes. *Oikos* **127**, 538–552. (doi:10.1111/oik.04527)

8. Cousineau SV, Alizon S. 2014 Parasite evolution in response to sex-based host heterogeneity in resistance and tolerance. *J. Evol. Biol.* **27**, 2753–2766. (doi:10.1111/jeb.12541)

9. Vale Pedro F., Choisy Marc, Little Tom J. 2013 Host nutrition alters the variance in parasite transmission potential. *Biol. Lett.* **9**, 20121145. (doi:10.1098/rsbl.2012.1145)

10. Shocket MS, Strauss AT, Hite JL, Šljivar M, Civitello DJ, Duffy MA, Cáceres CE, Hall SR. 2018 Temperature Drives Epidemics in a Zooplankton-Fungus Disease System: A Trait-Driven Approach Points to Transmission via Host Foraging. *Am. Nat.* **191**, 435–451. (doi:10.1086/696096)

11. Susi H, Vale PF, Laine A-L. 2015 Host Genotype and Coinfection Modify the Relationship of within and between Host Transmission. *Am. Nat.* **186**, 252–263. (doi:10.1086/682069)

12. Stephenson JF. 2019 Parasite-induced plasticity in host social behaviour depends on sex and susceptibility. *Biol. Lett.* **15**, 20190557. (doi:10.1098/rsbl.2019.0557)

13. Siva-Jothy JA, Vale PF. 2019 Viral infection causes sex-specific changes in fruit fly social aggregation behaviour. *Biol. Lett.* , 630913. (doi:10.1101/630913)

14. Keiser CN, Pinter-Wollman N, Augustine DA, Ziemba MJ, Hao L, Lawrence JG, Pruitt JN. 2016 Individual differences in boldness influence patterns of social interactions and the transmission of cuticular bacteria among group-mates. *Proc. R. Soc. B Biol. Sci.* **283**, 20160457–20160457. (doi:10.1098/rspb.2016.0457)

15. Susi H, Barrès B, Vale PF, Laine A-L. 2015 Co-infection alters population dynamics of infectious disease. *Nat. Commun.* **6**. (doi:10.1038/ncomms6975)

16. Lloyd-Smith JO, George D, Pepin KM, Pitzer VE, Pulliam JRC, Dobson AP, Hudson PJ, Grenfell BT. 2009 Epidemic dynamics at the human-animal interface. *Science* **326**, 1362–1367. (doi:10.1126/science.1177345)

17. Siva-Jothy JA, Vale PF. 2019 Dissecting genetic and sex-specific host heterogeneity in pathogen transmission potential. *bioRxiv* , 733915. (doi:10.1101/733915)

18. Kapun M, Nolte V, Flatt T, Schlötterer C. 2010 Host Range and Specificity of the Drosophila C Virus. *PLoS ONE* **5**, e12421. (doi:10.1371/journal.pone.0012421)

19. Webster CL *et al.* 2015 The Discovery, Distribution, and Evolution of Viruses Associated with Drosophila melanogaster. *PLOS Biol* **13**, e1002210. (doi:10.1371/journal.pbio.1002210)

20. Arnold PA, Johnson KN, White CR. 2013 Physiological and metabolic consequences of viral infection in Drosophila melanogaster. *J. Exp. Biol.* **216**, 3350–3357. (doi:10.1242/jeb.088138)

21. Chtarbanova S *et al.* 2014 Drosophila C virus systemic infection leads to intestinal obstruction. *J. Virol.* (doi:10.1128/JVI.02320-14)

22. Gupta V, Stewart CO, Rund SSC, Monteith K, Vale PF. 2017 Costs and benefits of sublethal Drosophila C virus infection. *J. Evol. Biol.* **30**, 1325–1335. (doi:10.1111/jeb.13096)

23. Siva-Jothy JA, Monteith KM, Vale PF. 2018 Navigating infection risk during oviposition and cannibalistic foraging in a holometabolous insect. *Behav. Ecol. Off. J. Int. Soc. Behav. Ecol.* **29**, 1426–1435. (doi:10.1093/beheco/ary106)

24. Anderson R, May R. 1991 *Infectious diseases of humans*. Oxford: Oxford University Press.

542 25. Siettos CI, Russo L. 2013 Mathematical modeling of infectious disease dynamics.
543 *Virulence* **4**, 295–306. (doi:10.4161/viru.24041)

544 26. Mackay TFC *et al.* 2012 The Drosophila melanogaster Genetic Reference Panel. *Nature*
545 **482**, 173–8. (doi:10.1038/nature10811)

546 27. Huszar T, Imler J. 2008 Drosophila Viruses and the Study of Antiviral Host-Defense. In
547 *Advances in Virus Research*, pp. 227–265. Academic Press.

548 28. Webster CL *et al.* 2015 The Discovery, Distribution, and Evolution of Viruses Associated
549 with Drosophila melanogaster. *PLoS Biol* **13**, e1002210–e1002210.
550 (doi:10.1371/journal.pbio.1002210)

551 29. Kapun M, Nolte V, Flatt T, Schlötterer C. 2010 Host Range and Specificity of the
552 Drosophila C Virus. *PLoS ONE* **5**. (doi:10.1371/journal.pone.0012421)

553 30. Csardi G, Nepusz T. 2006 The igraph software package for complex network research.
554 *InterJournal* **Complex Systems**, 1695.

555 31. Lloyd-Smith JO, Cross PC, Briggs CJ, Daugherty M, Getz WM, Latto J, Sanchez MS, Smith
556 AB, Swei A. 2005 Should we expect population thresholds for wildlife disease? *Trends*
557 *Ecol. Evol.* **20**, 511–519. (doi:10.1016/j.tree.2005.07.004)

558 32. Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ. 2007 Random
559 Forests for Classification in Ecology. *Ecology* **88**, 2783–2792. (doi:10.1890/07-0539.1)

560 33. Breiman L. 2001 Random forests. *Mach. Learn.* **45**, 5–32.
561 (doi:10.1023/A:1010933404324)

562 34. Ostfeld RS, Keesing F. 2012 Effects of Host Diversity on Infectious Disease. *Annu. Rev.*
563 *Ecol. Evol. Syst.* **43**, 157–182. (doi:10.1146/annurev-ecolsys-102710-145022)

564 35. Ferrari N, Cattadori IM, Nespereira J, Rizzoli A, Hudson PJ. 2004 The role of host sex in
565 parasite dynamics: field experiments on the yellow-necked mouse Apodemus flavicollis.
566 *Ecol. Lett.* **7**, 88–94. (doi:10.1046/j.1461-0248.2003.00552.x)

567 36. Grear DA, Perkins SE, Hudson PJ. 2009 Does elevated testosterone result in increased
568 exposure and transmission of parasites? *Ecol. Lett.* **12**, 528–537. (doi:10.1111/j.1461-
569 0248.2009.01306.x)

570 37. Zuk M, McKean KA. 1996 Sex differences in parasite infections: Patterns and processes.
571 *Int. J. Parasitol.* **26**, 1009–1024. (doi:10.1016/S0020-7519(96)80001-4)

572 38. Baxter CM, Barnett R, Dukas R. 2015 Aggression, mate guarding and fitness in male fruit
573 flies. *Anim. Behav.* **109**, 235–241. (doi:10.1016/j.anbehav.2015.08.023)

574 39. Craft ME, Caillaud D. 2011 Network Models: An Underutilized Tool in Wildlife
575 Epidemiology? *Interdiscip. Perspect. Infect. Dis.* (doi:10.1155/2011/676949)

40. Keiser CN, Pinter-Wollman N, Ziemba MJ, Kothamasu KS, Pruitt JN. 2017 The index case is not enough: Variation among individuals, groups and social networks modify bacterial transmission dynamics. *J. Anim. Ecol.* , 1–10. (doi:10.1111/1365-2656.12729)

41. Keiser CN, Howell KA, Pinter-Wollman N, Pruitt JN. 2016 Personality composition alters the transmission of cuticular bacteria in social groups. *Biol. Lett.* **12**, 20160297. (doi:10.1098/rsbl.2016.0297)

42. Barron DG, Gervasi SS, Pruitt JN, Martin LB. 2015 Behavioral competence: How host behaviors can interact to influence parasite transmission risk. *Curr. Opin. Behav. Sci.* **6**, 35–40. (doi:10.1016/j.cobeha.2015.08.002)

43. Keiser CN, Rudolf VHW, Sartain E, Every ER, Saltz JB. 2018 Social context alters host behavior and infection risk. *Behav. Ecol.* **29**, 869–875. (doi:10.1093/beheco/ary060)

44. Nunn CL, Jordán F, McCabe CM, Verdolin JL, Fewell JH. 2015 Infectious disease and group size: more than just a numbers game. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**.

45. Sah P, Leu ST, Cross PC, Hudson PJ, Bansal S. 2017 Unraveling the disease consequences and mechanisms of modular structure in animal social networks. *Proc. Natl. Acad. Sci.* **114**, 4165–4170. (doi:10.1073/pnas.1613616114)

46. Wallace R, Wallace RG. 2015 Blowback: new formal perspectives on agriculturally driven pathogen evolution and spread. *Epidemiol. Infect.* **143**, 2068–2080. (doi:10.1017/S0950268814000077)

47. Siva-Jothy JA, White LA, Craft ME, Vale PF. 2019 Population-Level Disease Dynamics Reflect Individual Heterogeneities in Transmission. *bioRxiv* , 735480. (doi:10.1101/735480)

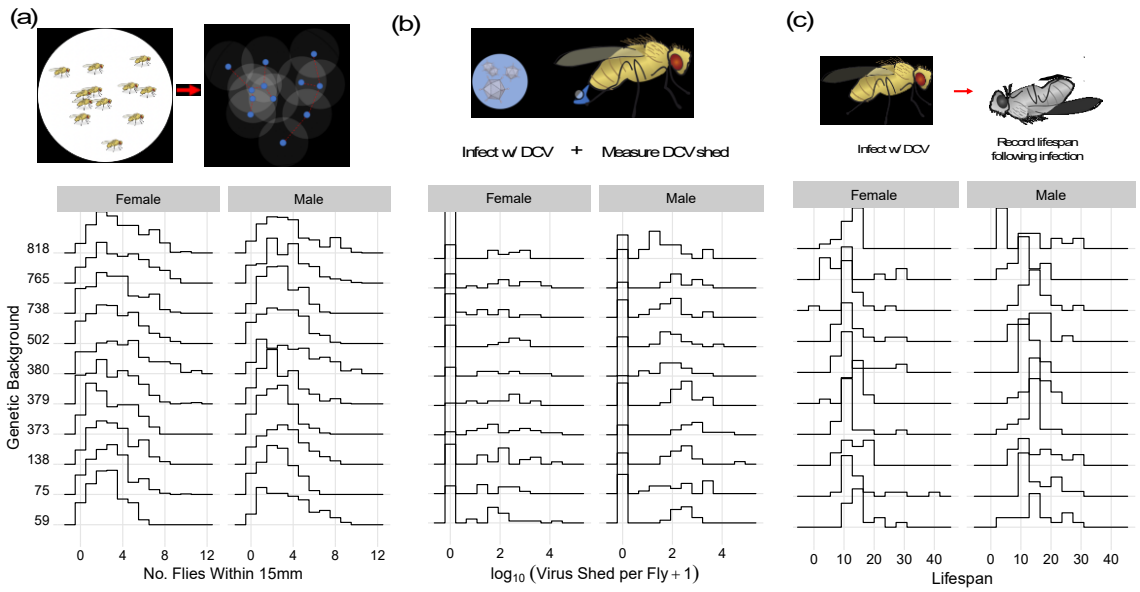

Figure 1 – The epidemiological model was parameterised by sampling frequency distributions of experimental data collected from female and male ten Drosophila Genetic Reference Panel lines infected with DCV published previously [13,17]. Here, we provide a qualitative description of these data. a) social aggregation: the average number of neighbouring flies present within a 15mm radius of each focal fly; b) infectiousness: the number of viral copies shed per fly within the first 3 days following infected, as measured by DCV-specific qPCR; c) the day of death of each individual infected fly. Detailed analysis showing extensive line-by-sex interactive effects are reported in [13,17].

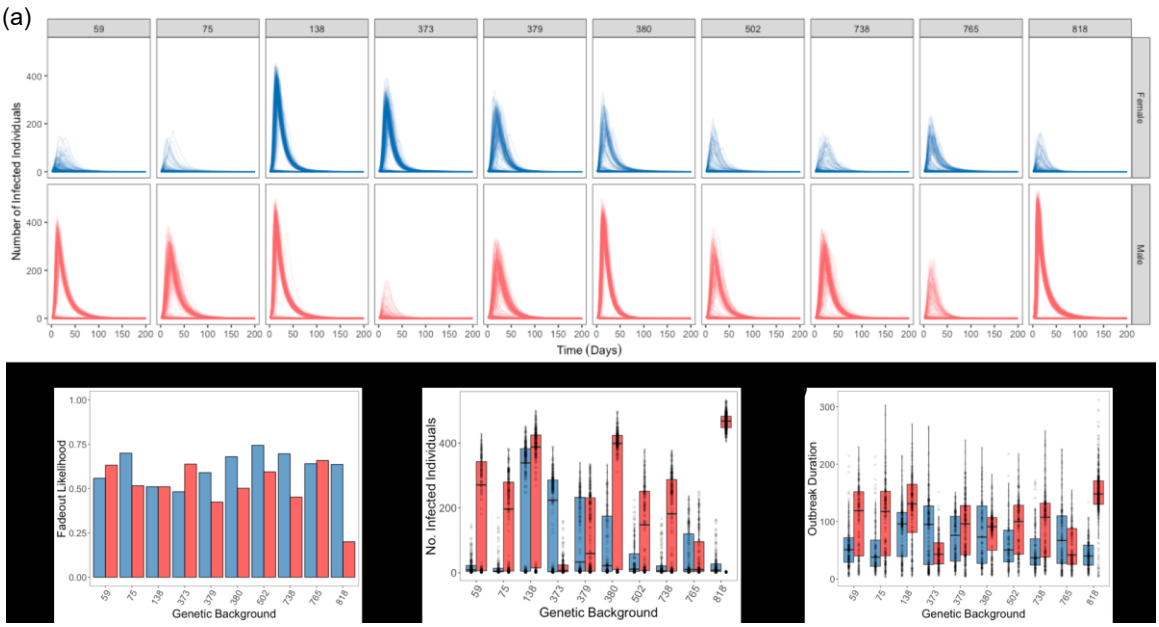

Figure 2 – a) Simulation time courses of populations comprised of either male (red) or female (blue) individuals of the same sex and genetic background (columns) for simulation experiment #1. Across all of these simulations, parameters outside of host genetic background and sex are fixed; threshold radius (r)= 15mm, transmission efficiency (τ)=1 and scaled infectiousness (η)=2. **(b-d)** Summary statistics of simulations of populations comprised of male (red) or female (blue) individuals of the same genetic background (x-axis) for (b) the proportion of simulations that resulted in fadeout; and, in the subset of simulations where fadeout did not occur and disease spread from the index case; (c) the maximum number of infected individuals at any given time step; and (d) the number of time steps infected by the index case. Shown for threshold radius (r)= 15mm, transmission efficiency (τ)=1 and scaled infectiousness (η)=2. A random forest analysis was used to determine the relative importance of genetic background and sex to each summary statistic used to describe outbreak dynamics (Figure S1).

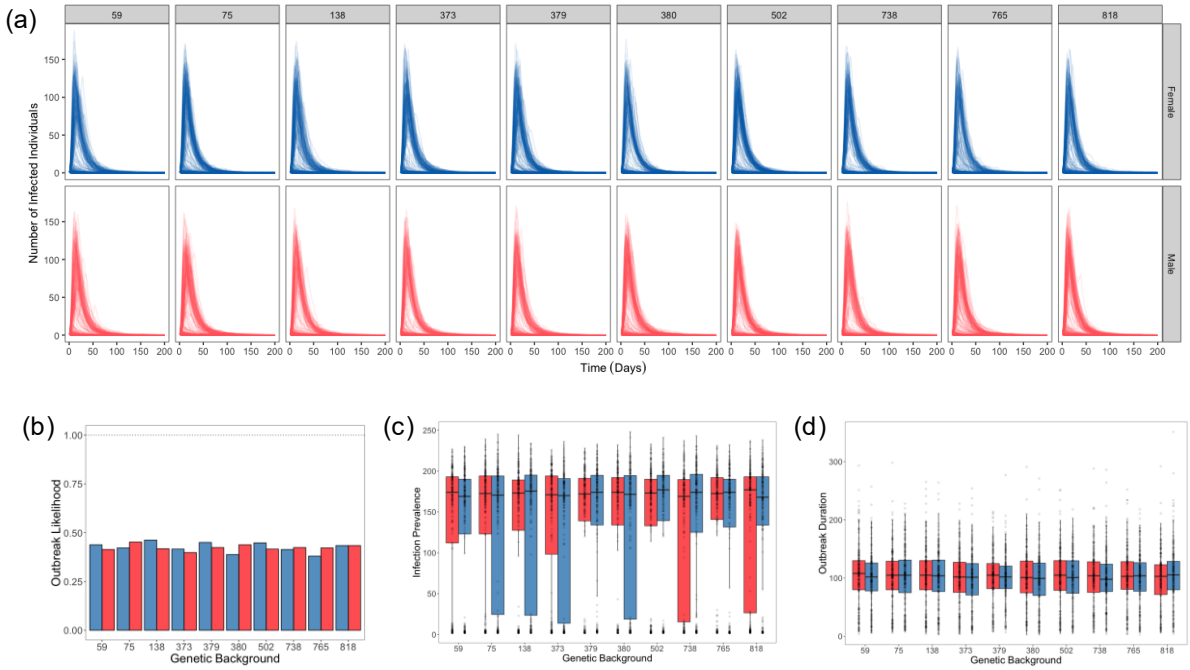

Figure 3 – Simulation time courses of populations comprised of all ten genetic backgrounds and males (red), and females (blue) in equal proportion, where the index case of an outbreak is an individual of a specific genetic background and sex (simulation experiment #2). Across all of these simulations, other parameters are fixed: threshold radius (r)= 15mm, transmission efficiency (τ)=1 and scaled infectiousness (η)=2. A random forest analysis was used to determine the relative importance of genetic background and sex to each summary statistic used to describe outbreak dynamics (Figure S2).

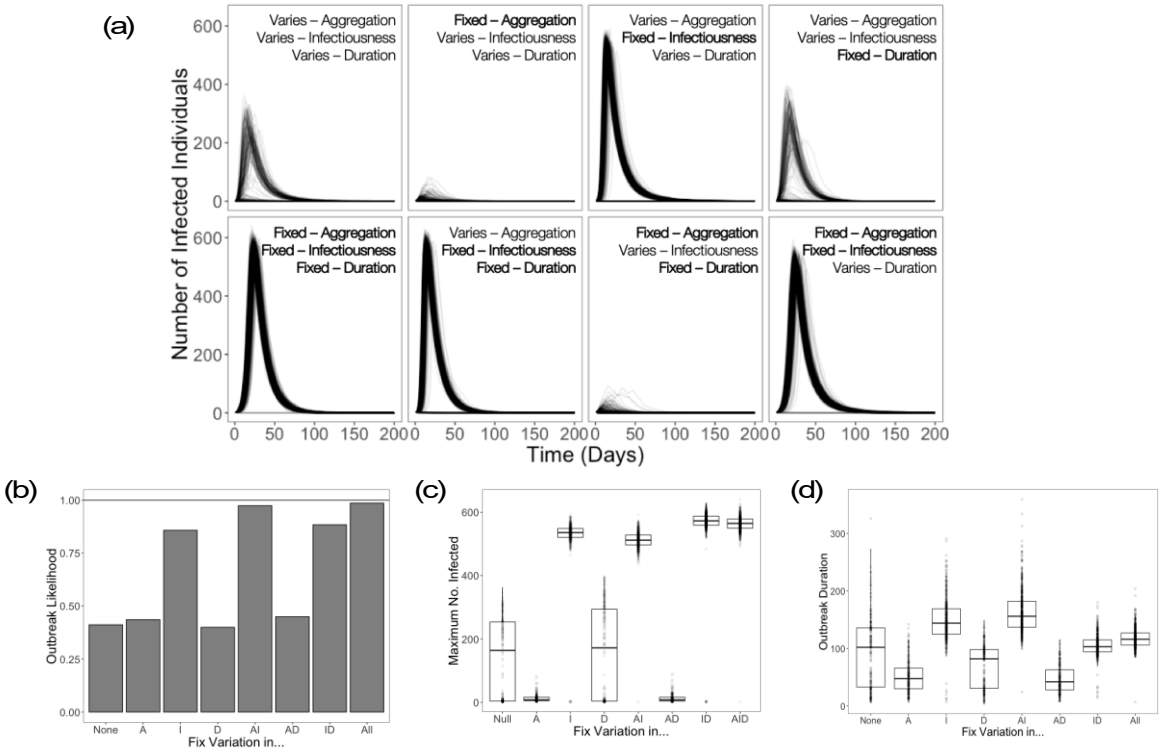

Figure 4 – a) Simulation time courses of populations where aggregation, infectiousness and duration variation are derived from the entire population's variation rather than for a single genetic line and sex combination (simulation experiment #3). In each panel, the variation of a particular set of components is confined to the population's mean. Across all of these simulations, parameters outside of host genetic background and sex are fixed: threshold radius = 15mm, transmission efficiency =1 and scaled infectiousness =2. **(b-d)** Summary statistics of time course simulations where individual variation is determined by the variation seen across all genetic backgrounds and sexes (simulation experiment #3). The x-axis of all panels sees variation in aggregation (A), infectiousness (I) and mortality rate (D), and all their combinations fixed to the population mean. Outbreak metrics include: b) the proportion of simulations that resulted in fadeout; c) the maximum number of individuals infected during the simulation; and d) the time until maximum prevalence was reached. Shown for threshold radius = 15mm, transmission efficiency =1, and scaled infectiousness =2. A random forest analysis was used to determine the relative importance of genetic background and sex to each summary statistic used to describe outbreak dynamics (Figure S3).

Parameter	Levels	Simulation 1	Simulation 2	Simulation 3
Population genetic background	RAL-59, RAL-75, RAL-138, RAL-373, RAL-379, RAL-380, RAL-502, RAL-738, RAL-765, RAL-818	X		
Population sex	Female, Male	X		
Index genetic background	RAL-59, RAL-75, RAL-138, RAL-373, RAL-379, RAL-380, RAL-502, RAL-738, RAL-765, RAL-818		X	
Index sex	Female, Male		X	
Threshold radius (r)	10mm, 15mm, 20mm	X	X	X
Pathogen transmission efficiency (τ)	0.1, 0.5, 1	X	X	X
Scaled infectiousness (η)	1, 2	X	X	X
Vary social aggregation	TRUE, FALSE			X
Vary infectiousness	TRUE, FALSE			X
Vary infection duration	TRUE, FALSE			X

Table 1. Parameters used to simulate outbreaks of infectious disease in simulations 1-3. Simulation 1 tested the effect of genetic and sex-specific variation in social aggregation, viral shedding and susceptibility on population-level disease dynamics. Simulation 2 tested the effect of susceptible host diversity on disease transmission potential. Simulation 3 tested the effect of variation in social aggregation, infectiousness and infection duration on population-level disease transmission dynamics. We conducted 500 replicates per parameter set with 1000 individuals in the network. Simulations were allowed to run for 1000-time steps.