Bridging the gap between theory and data: the Red Queen Hypothesis for sex

Sang Woo Park and Benjamin M Bolker

May 4, 2019

Abstract

Sexual reproduction continues to persist in nature despite its large cost. The Red Queen Hypothesis for sex postulates that sexual reproduction is maintained in the host population under strong parasitic pressure by producing genetically rare offspring that are resistant to infection. Mathematical models have been used to lay theoretical foundations for the hypothesis and empirical data to confirm model predictions. For example, Lively used a simple host-parasite model to predict that the frequency of sexual hosts should be positively correlated with the prevalence of infection; Lively later confirmed the prediction through numerous field studies of snail-trematode systems in New Zealand. In this study, we fit a simple metapopulation host-parasite coevolution model to three data sets, each representing a different snail-trematode system, by matching the observed prevalence of sexual reproduction and trematode infection among hosts. Using the inferred parameters, we perform a power analysis to test how likely it is to observe the positive correlation, predicted by Lively. We discuss ways in which our model prediction differs from empirical data and provide practical guidance to both modelers and empiricists. Overall, our study suggests that a simple Red Queen model can only partially explain the observed relationships between parasite infection and the maintenance of sexual reproduction.

1 Introduction

Despite being the dominant mode of reproduction in multicellular organisms (Vrijenhoek, 1998), sexual reproduction entails numerous costs (Lehtonen et al., 2012). The most commonly mentioned is the cost of producing males (Smith, 1978): as males cannot produce offspring, sexual lineages are expected to be outgrown by their asexual counterparts that can grow as twice as fast. This infamous *two-fold cost of sex* (Smith, 1978) relies on the assumption that everything else is equal. Then, what else is not equal and drives the sex to persist?

One explanation for the persistence of sexual reproduction is the Red Queen Hypothesis (Bell, 1982). The Red Queen Hypothesis for sex predicts that sexu-

ally reproducing hosts overcome the cost of sex under strong parasitic pressure by producing genetically diverse offspring that can escape infection (Haldane, 1949; Jaenike, 1978; Hamilton, 1980; Hamilton et al., 1990). Limited genetic diversity of asexually reproducting hosts incurs a greater fitness cost to infection, allowing for sexual reproduction (or, more precisely, greater genetic diversity (King et al., 2011)) to be maintained in the host population (Ashby and King, 2015).

Much of the theoretical literature has focused on determining qualitative conditions under which parasite selection can maintain sexual reproduction in the host population. Here, we describe a few important ones. First, hosts and parasites must coevolve (Bell, 1982). Host-parasite coevolution creates a time-lagged selective advantage for rare host genotypes, creating an oscillation in genotypic frequencies and allowing for sexual reproduction to be maintained in the host population (Clarke, 1976; Jaenike, 1978; Hamilton, 1980; Agrawal and Lively, 2001). Second, parasites must be highly virulent (May and Anderson, 1983). Although sexual and asexual hosts can coexist at intermediate virulence (Howard et al., 1994), sexual reproduction is unable to provide enough advantage to overcome the cost of sex against avirulent parasites (Howard et al., 1994). Finally, sexual hosts must be genetically more diverse than asexual hosts, as high clonal diversity may diminish the advantage of sexual reproduction (Lively and Howard, 1994; Lively, 2010c; Ashby and King, 2015).

Some theoretical studies have departed from the classical population genetics framework to study effects of ecological and epidemiological structures on the Red Queen dynamics. A few studies suggested that incorporating more realistic ecological and epidemiological details may assist in supporting sexual reproduction in the host population (Lively, 2009, 2010b) and maintaining coevolutionary cycle (Ashby and Gupta, 2014). In contrast to these findings, MacPherson and Otto (2018) showed that Red Queen dynamics (i.e., cycles in allele frequencies) fail to persist when explicit epidemiological structure is taken into account with coevolutionary dynamics. Given a wide range of assumptions that these eco-evolutionary models make, it may be difficult to understand how different ecological details of these models affect the overall maintenance of sexual reproduction; however, what is clear is that these details play important roles in shaping the coevolutionary dynamics (Song et al., 2015; Ashby et al., 2019; Haafke et al., 2016).

On the other hand, empirical studies have mostly focused on confirming predictions that stem from the Red Queen Hypothesis. Typical among them are rapid parasite evolution (Rauch et al., 2006), local adaptation (Lively, 1989; Morran et al., 2014; King et al., 2011; Gibson et al., 2016), time-lagged selection (Buckling and Rainey, 2002; Decaestecker et al., 2007; Koskella and Lively, 2009; Thrall et al., 2012; Koskella, 2013, 2014), and association between parasite prevalence and host reproductive mode (Lively, 1992; Vergara et al., 2013; Verhoeven and Biere, 2013). A key example is the snail population in New Zealand that serves as an intermediate hosts for trematodes (Winterbourn, 1974; McArthur and Featherston, 1976). Through a few decades of work, Lively *et al.* have demonstrated that the population satisfies necessary conditions for the

host-parasite coevolutionary dynamics and provides support for the hypothesis (e.g., Lively (1987, 1989); Dybdahl and Lively (1995b, 1998); Jokela et al. (2009); Vergara et al. (2014b); Gibson et al. (2016)). While field studies provide only indirect evidence for the hypothesis, recent studies show that more direct evidence can be achieved using experimental systems (Auld et al., 2016; Slowinski et al., 2016).

Even though the Red Queen Hypothesis has gained both theoretical and empirical support, there still remains a gap between theory and data. Many theoretical models rely on simplifying assumptions that are not applicable to natural populations and make indirect connections to empirical studies. For example, none of the Red Queen models reviewed by Ashby and King (2015) make statistical connections to empirical data. It is unclear how well these models capture coevolutionary dynamics in natural systems.

However, theoretical models can still be used to make *qualitative* predictions about nature. Lively (1992) initially postulated that infection prevalence should be positively correlated with frequency of sexual hosts and later formalized the idea using a mathematical model (Lively, 2001). The prediction has since been confirmed in several empirical studies, most of which are based on the snail-trematode system (Lively and Jokela, 2002; Kumpulainen et al., 2004; King et al., 2011; Vergara et al., 2013; McKone et al., 2016; Gibson et al., 2016). Surprisingly, such correlation was not observed in a different snail-trematode system (Heller and Farstey, 1990; Ben-Ami and Heller, 2005, 2007, 2008; Dagan et al., 2013a).

Here, we try to help bridge the gap between theory and data in understanding the role of parasites in maintaining sexual reproduction in host populations and make quantitative inference about empirical systems. We extend the model used by Lively (2010b) to account for demographic stochasticity and include a simple metapopulation structure. Then, we fit the model to data sets from three field studies (Dagan et al., 2013a; McKone et al., 2016; Vergara et al., 2014b) using Approximate Bayesian Computation (ABC) to estimate biologically relevant parameters. We assess model fits and discuss discrepancies between a theoretical model and the observed data. Using biologically realistic parameters, we test for power (probability of observing a significant effect) to detect a positive correlation bdetween frequency of sexual reproduction and prevalence of infection Lively (2001). By performing power analysis, we provide guidance for studying the Red Queen Hypothesis and discuss underlying factors that drive the correlation.

2 Methods

2.1 Data

We analyze three observational data sets (Vergara et al., 2014b; McKone et al., 2016; Dagan et al., 2013a) that represent interactions between freshwater snails and sterilizing trematodes. The New Zealand snail populations (Vergara et al.,

2014b; McKone et al., 2016) consist of *Potamopyrgus antipodarum*, which serves as an intermediate host for trematodes, among which *Microphallus* sp. is most commonly found (Winterbourn, 1974; Lively, 1987). The Israeli snail populations (Dagan et al., 2013a) consist of *Melanoides tuberculata*, which also serves as an intermediate host for trematodes, including *Centrocestus* sp. (Ben-Ami and Heller, 2005), *Philophthalmus* sp. (Ben-Ami, 2006), and *Transversotrema patialense* (Ben-Ami et al., 2005). Both the New Zealand (Phillips and Lambert, 1989; Wallace, 1992; Dybdahl and Lively, 1995a) and the Israeli (Samadi et al., 1999) snail populations are able to maintain coexistence of diploid sexual females and polyploid asexual females.

We choose to analyze snail-trematode systems because they have been extensively studied under the context of the Red Queen Hypothesis. The snail-trematode systems in New Zealand have been shown to exhibit positive correlation between prevalence of infection and frequency of sexual hosts (Lively and Jokela, 2002; King et al., 2011; Vergara et al., 2013; McKone et al., 2016; Gibson et al., 2016), local adaptation (Lively, 1989; Lively et al., 2004; King et al., 2011), negative frequency-dependent selection (Dybdahl and Lively, 1995b, 1998; Jokela et al., 2009; Koskella and Lively, 2009), and periodic selection for sexual reproduction (Vergara et al., 2014b; Gibson et al., 2018); these results are consistent with the host-parasite coevolutionary theory behind the Red Queen Hypothesis. Therefore, we expect a basic Red Queen model to be able to mimic the observed dynamics reasonably well (compared to other systems that may support the hypothesis but are not as well understood).

We included the Israeli population because there has been a consistent lack of positive correlation between the prevalence of infection and frequency of sexual hosts (Heller and Farstey, 1990; Ben-Ami and Heller, 2005, 2007, 2008; Dagan et al., 2013a), which sharply contrasts the observations from the New Zealand population. Comparison of these two populations with a mathematical model may allow us to better understand the underlying cause of the difference.

The data sets include proportional of infected snails, proportion of sexual/asexual snails, and locations of sampling sites; the data set from Vergara et al. (2014b) also includes sampling year, which spans 5 years. Whereas Vergara et al. (2014b) report infection status of snails specific to *Microphallus* sp., McKone et al. (2016) and Dagan et al. (2013a) do not distinguish among different spcies. The data sets collected by Dagan et al. (2013a) and Vergara et al. (2014b) are obtained from their Dryad repositories (Dagan et al., 2013b; Vergara et al., 2014a). The data set collected by McKone et al. (2016) is extracted from their figure.

2.2 Model

We model obligately sexual hosts competing with obligately asexual hosts in a metapopulation by extending the model introduced by Lively (2010b). Our model is a discrete time susceptible-infected (SI) model with natural mortality and virulence (defined as reduction in offspring production among infected hosts). Metapopulation structure is included to model unobserved dynamics

among different habitats; equivalently, each population can be considered as a sampling site. A similar metapopulation model was developed by Lively (2018) in order to study local adaptaion of parasites.

We do not model the life history of snail-trematode interactions explicitly. Modeling interactions of multiple trematodes species, each of which go through a different life cycle, with a single snail species is extremly complicated. Instead, testing a basic model against real data will allow us to identify the difference between theory and data more clearly.

All hosts are assumed to be diploids with two biallelic loci controlling resistance, and parasites are assumed to be haploids. Let $S_{ij}^k(t)$ and $A_{ij}^k(t)$ be the number of sexual and asexual hosts with genotype ij from population k at generation t, where each subscript, i and j, represents a host hapolotype: $i, j \in \{\mathrm{AB}, \mathrm{Ab}, \mathrm{aB}, \mathrm{ab}\}$. For simplicity, we drop the superscript representing population and write $S_{ij}(t)$ and $A_{ij}(t)$; every population is governed by the same set of equations unless noted otherwise (e.g., when we account for interaction between populations). Following Lively (2010b), the expected genotypic contribution (before recombination or outcrossing) by sexual hosts is given by

$$S'_{ij} = c_b(1-s)\bigg(W_U S_{ij,U}(t) + W_I S_{ij,I}(t)\bigg), \tag{1}$$

where s is the proportion of males produced by sexual hosts, and $S_{ij,U}$ and $S_{ij,I}$ are the number of uninfected and infected sexual hosts in a population. W_U and W_I represent their corresponding fitnesses where virulence is defined as $V = 1 - W_I/W_U$. We allow for the cost of sex to vary by multiplying a scale parameter, c_b, to the growth term, where $2/c_b$ corresponds to a two fold cost of sex when $c_b = 1$ (Ashby and King, 2015). Recombination and outcrossing are modeled after incorporating genotypic contributions from other populations.

We define

$$W_U = \frac{b_U}{1 + a_U N(t)}, \; W_I = \frac{b_I}{1 + a_I N(t)}$$

where b_U and b_I are number of offspring produced by uninfected and infected hosts, respectively, and a_U and a_I determine their corresponding strengths of density dependence (Lively, 2010b; Smith and Slatkin, 1973). For simplicity, we assume that $a_U = a_I$ so that virulence can be defined strictly in terms of decrease in offspring production and is density independent: $V = 1 - b_I/b_U$.

Asexual hosts are assumed to be strictly clonal. Then, the expected genotypic contribution by asexual hosts is given by

$$A'_{ij} = W_U A_{ij,U}(t) + W_I A_{ij,I}(t), \tag{2}$$

where $A_{ij,U}$ and $A_{ij,I}$ are the number of uninfected and infected asexual hosts in a population.

We assume that a proportion ϵ_{mix} of a population mixes with other populations. Then, the expected number of offspring in the next generation (accounting

for contributions from all populations) is given by

$$\begin{aligned}
\mathrm{E}\left(S_{ij}^{k}(t+1)\right) &= f_{\mathrm{sex}}\left((1-\epsilon_{\mathrm{mix}})\left(S_{ij}^{k}\right)' + \frac{\epsilon_{\mathrm{mix}}}{n_{\mathrm{pop}}-1}\sum_{h\neq k}\left(S_{ij}^{h}\right)'\right), \\
\mathrm{E}\left(A_{ij}^{k}(t+1)\right) &= (1-\epsilon_{\mathrm{mix}})\left(A_{ij}^{k}\right)' + \frac{\epsilon_{\mathrm{mix}}}{n_{\mathrm{pop}}-1}\sum_{h\neq k}\left(A_{ij}^{h}\right)',
\end{aligned} \tag{3}$$

where $f_{\mathrm{sex}}(x)$ is the function that models sexual reproduction, including recombination probability r_{host} and outcrossing, and n_{pop} is the number of populations modeled. Then, the total number of sexual and asexual hosts in the next generation given by Poisson random variables with mean specified previously. We also allow for stochastic migration in order to avoid fixation:

$$\begin{aligned}
S_{ij}^{k}(t+1) &\sim \mathrm{Poisson}\left(\lambda = \mathrm{E}\left(S_{ij}^{k}(t+1)\right)\right) + \mathrm{Bernoulli}\left(p = p_{ij,\mathrm{sex}}\right), \\
A_{ij}^{k}(t+1) &\sim \mathrm{Poisson}\left(\lambda = \mathrm{E}\left(A_{ij}^{k}(t+1)\right)\right) + \mathrm{Bernoulli}\left(p = p_{ij,\mathrm{asex}}\right),
\end{aligned} \tag{4}$$

where $p_{ij,\mathrm{sex}}$ and $p_{ij,\mathrm{asex}}$ are the probabilities of a sexual and an asexual host with genotype ij entering a popualation.

Infection is modeled using the matching alleles model (Otto and Michalakis, 1998). We assume that snails are equally susceptible to parasites that match either haplotype. However, parasites must match the host haplotype at both loci in order to successfully infect a host. The expected number of infected hosts that is infected with parasite with genotype i at generation t is given by:

$$I_i(t) = f_{\mathrm{mutation}}\left(\sum_j \left(S_{ij,i,I}(t) + A_{ij,i,I}(t)\right)\right). \tag{5}$$

$S_{ij,i,I}(t)$ and $A_{ij,i,I}(t)$ represent the expected numbers of sexual and asexual hosts that have genotype ij and are infected with parasites that have genotype i. The function f_{mutation} models mutation at a single locus with with probability r_{parasite} (Ashby and King, 2015).

The total expected number of infectious contacts made by infected hosts within a population is given by $\lambda_i^k = \beta^k {I_i'}^k(t)$, where β^k is the transmission rate of each population, and $I_i'(t)$ is the number of infected hosts accounting for stochastic migration with probability $p_{i,\mathrm{parasite}}$ to avoid fixation. We model mixing between subpopulations by allowing infected hosts to make contact with susceptible hosts in other populations; we assume that proportion ϵ_{mix} of the infectious contacts are equally distributed among other subpopulations. Then, the total amount of infectious contact, coming from hosts that carry genotype i parasite, that is received by susceptible hosts in population k is given by

$$\lambda_{i,\mathrm{total}}^k = (1-\epsilon_{\mathrm{mix}})\lambda_i^k + \frac{\epsilon_{\mathrm{mix}}}{n_{\mathrm{pop}}-1}\sum_{l\neq k}\lambda_i^l \tag{6}$$

Then, the force of infection that a susceptible host with genotype ij experiences in generation $t+1$ is given by

$$\mathrm{FOI}_{ij}^{k} = \frac{\lambda_{i,\mathrm{total}}^{k} + \lambda_{j,\mathrm{total}}^{k}}{2N^{k}(t+1)}, \tag{7}$$

where $N^{k}(t+1) = \sum_{i,j} S_{ij}^{k}(t+1) + A_{ij}^{k}(t+1)$ is the total number of hosts in generation $t+1$. The probability that a susceptible host with genotype ij in population k becomes infected in the next generation is given by

$$P_{ij}^{k}(t+1) = 1 - \exp\left(\mathrm{FOI}_{ij}^{k}\right). \tag{8}$$

Finally, the number of infected hosts in the next generation is determined by a binomial process:

$$\begin{aligned} S_{ij,I}^{k}(t+1) &\sim \mathrm{Binom}\left(S_{ij}^{k}(t+1), P_{ij}^{k}(t+1)\right), \\ A_{ij,I}^{k}(t+1) &\sim \mathrm{Binom}\left(A_{ij}^{k}(t+1), P_{ij}^{k}(t+1)\right). \end{aligned} \tag{9}$$

The expected number of infected hosts that have genotype ij and are infected by parasites with genotype i in the next generation is proportional to the amount of infectious contact that was made in the current generation:

$$\begin{aligned} S_{ij,i,I}^{k}(t+1) &= \frac{2^{\delta_{ij}}\lambda_{i,\mathrm{total}}^{k}}{\lambda_{i,\mathrm{total}}^{k} + \lambda_{j,\mathrm{total}}^{k}} S_{ij,I}^{k}(t+1) \\ A_{ij,i,I}^{k}(t+1) &= \frac{2^{\delta_{ij}}\lambda_{i,\mathrm{total}}^{k}}{\lambda_{i,\mathrm{total}}^{k} + \lambda_{j,\mathrm{total}}^{k}} A_{ij,I}^{k}(t+1) \end{aligned} \tag{10}$$

where a δ_{ij} is a Kronecker delta (δ_{ij} equals 1 when $i = j$ and 0 otherwise).

2.3 Simulation design and parameterization

Many Red Queen models have focused on competition between a single asexual genotype and multiple sexual genotypes or have assumed equal genetic diversity between asexual and sexual hosts (see (Ashby and King, 2015) for a review of previous Red Queen models) but neither of these assumptions is realistic. Instead, Ashby and King (2015) adopted a more realistic approach by incorporating stochastic external migration of an asexual genotype to a population and allowing for asexual genetic diversity to vary over time. Here, we combine these methods. We allow for stochastic external migration of asexual hosts with different genotypes into the system but fix the number of asexual genotypes (denoted by G_{asex}) that can be present in the system. Since sexual hosts have limited genetic diversity in our model (diploid hosts with two biallelic loci yields a total of 10 genotypes), allowing for unlimited migration of asexual hosts will cause the sexual population to be easily outcompeted by the asexual population. Limiting asexual genetic diversity allows us to account for the intrinsic

difference in sexual and asexual diveristy and make a compromise between simple and realistic models. At the beginning of each simulation, a pool of G_{asex} asexual genotypes are sampled at random the entire genotypic space; these are the genotypes that are available for external immigration into subpopulations. Then, we use our model-based comparison with data to estimate G_{asex} to test whether greater asexual genetic diversity can be supported than a typically assumed one-to-many ratio.

To account for differing number of sexual and asexual genotypes, we let

$$\begin{aligned} p_{ij,\text{sex}} &= 1-(1-p_{\text{host}})^{1/G_{\text{sex}}}, \\ p_{ij,\text{asex}} &= \begin{cases} 1-(1-p_{\text{host}})^{1/G_{\text{asex}}} & \text{if } ij \in \{\text{asexual genotypes}\} \\ 0 & \text{otherwise} \end{cases}, \end{aligned} \tag{11}$$

where p_{host} is the probability that at least one sexual or asexual host enters the population in a generation. We scale the probability of an infected host carrying parasite genotype i in a similar way for interpretability:

$$p_{i,\text{parasite}} = 1-(1-p_{\text{infected}})^{1/G_{\text{parasite}}}, \tag{12}$$

where p_{infected} is the probability that at least one infected host enters the population in a generation. The number of parasite genotypes G_{parasite} is equal to 4 (because parasites are assumed to be haploids with two loci).

Each simulation consists of 40 subpopulations. Every subpopulation is initialized with 2000 sexual hosts, of which 80 are infected. They are assumed to be in Hardy-Weinberg equilibrium where allele frequencies in each locus is half. The transmission rate, β^k, is randomly drawn for each subpopulation from a Gamma distribution with mean β_{mean} and coefficient of variation β_{CV}. Simulation runs for 500 generations without introduction of asexuals. At generation 501, 10 asexual hosts of a single genotype are introduced to each population (the asexual genotype introduced can vary across population) and simulation runs for 600 generations while allowing for stochastic migration of asexuals.

2.4 Approximate Bayesian Computation

We use Approximate Bayesian Computation (ABC) to estimate parameters from data (Toni et al., 2009). ABC relies on comparing summary statistics of observed data and those of simulated data and is particularly useful when the exact likelihood function is not available. We consider mean proportion of infected and sexually reproducing snails in the system and variation in these proportions – measured by coefficient of variation (CV) – across space (population) and time as our focal summary statistics. These summary statistics are calculated for both observed and simulated data and are used in ABC. As Dagan et al. (2013a) and McKone et al. (2016) only reported the proportion of males, the proportion of sexual hosts is assumed to be twice the proportion of males.

CV across space is calculated by first calculating mean proportions by averaging across time (generation) for each site (subpopulation) and then taking the

the CV of these mean proportions. CV across time (generation) is calculated by first averaging proportions across space (subpopulation) at each generation and then taking the CV. For purely spatial data (Dagan et al. (2013a) and McKone et al. (2016)), CV across space is calculated without averaging across time.

Because ABC is a Bayesian method, we want to specify prior distributions for all parameters. We use weakly informative priors for all parameters that we estimate except c_b, a scale parameter for the cost of sex (see Table 1 for prior distributions used and parameters assumed). The prior distribution for the scale parameter is chosen so that 95% prior quantile of cost of sex ($2/c_b$) is approximately equal to the 95% confidence interval reported by Gibson et al. (2017). For brevity, all other parameters are assumed to be fixed. While it is a common practice to fix several parameters of a Red Queen model (e.g., (Lively, 2010b; Ashby and King, 2015; Haafke et al., 2016; Ashby et al., 2019)), not explicitly accounting for uncertainty in parameters of a model can lead to overly confident conclusions (Elderd et al., 2006).

We use the Population Monte Carlo approach (Turner and Van Zandt, 2012), which allows for sampling more efficiently while ensuring that final result still satisfies criteria to be a correct (approximate) Bayesian posterior. The Population Monte Carlo approach begins with a basic ABC. For each random parameter sample drawn from the prior distribution, the model is simulated and a sample of subpopulation is drawn from the simulated system, equal to the number of sites collected in a study. Then, summary statistics are calculated based on the last 100 generations out of 1100 generations; current parameter set is accepted if the distance between simulated and observed data is less than a specified tolerance value. Distance is measured by the sum of absolute differences in summary statistics between simulated and observed data. This process is repeated until 100 parameter sets are accepted.

After the first run ($t = 1$), equal weights ($w_{i,1} = 1/100$) are assigned to each accepted parameter set $\boldsymbol{\theta}_{i,1}$, where $1 \leq i \leq 100$. a weighted random sample ($\boldsymbol{\theta}^*$) is drawn from the accepted parameters of the previous run ($t - 1$) with weights $w_{i,t-1}$ and For any run $t > 1$, a parameter sample ($\boldsymbol{\theta}_{i,t}$) is proposed from a multivariate normal distribution with a mean $\boldsymbol{\theta}^*$ and a variance covariance matrix that is equal to $\sigma^2_{t-1} = 2\text{Var}(\boldsymbol{\theta}_{1:N,t-1})$, where $\text{Var}(\boldsymbol{\theta}_{1:N,t-1})$ is the weighted variance covariance matrix of the accepted parameters from the previous run. N is the total number of accepted parameters from the previous run.

G_{asex} is rounded to the nearest integer and the model is simulated. If a proposed parameter is accepted, the following weight is assigned:

$$w_{i,t} = \frac{\pi(\boldsymbol{\theta}_{i,t})}{\sum_{i=1}^{100} w_{j,t-1} q(\boldsymbol{\theta}_{j,t-1} | \boldsymbol{\theta}_{i,t}, \sigma^2_{t-1})}$$

where $\pi(\cdot)$ is a prior density and $q(\cdot|\boldsymbol{\theta}_{i,t}, \sigma^2_{t-1})$ is a multivariate normal density with mean $\boldsymbol{\theta}_{i,t}$ and variance covariance matrix σ^2_{t-1}. For each run, 100 parameters are accepted and weights are normalized at the end to sum to 1.

For each observed datum, we perform 4 runs with decreasing toleranc. First

Notation	Description	Prior distribution/parameter values	Source
β_{mean}	Mean transmission rate	Gamma($k = 2, \theta = 10$)	Assumption
β_{CV}	CV transmission rate	Gamma($k = 2, \theta = 0.5$)	Assumption
V	Virulence	Beta($\alpha = 6, \beta = 2$)	Assumption
ϵ_{mix}	Mixing proportion	Beta($\alpha = 1, \beta = 9$)	Assumption
$G_{\text{asex}} - 1$	Number of asexual genotypes - 1	BetaBinomial($N = 9, p = 3/9, \theta = 5$)	Assumption
c_b	Cost of sex scale	LogNormal($\mu = -0.07, \sigma = 0.09$)	Gibson et al. (2017)
s	Proportion of male offspring produced	0.5	Assumption
b_U	Number of offspring produced by an uninfected host	20	Lively (2010b)
b_I	Number of offspring produced by an infected host	$(1 - V)b_U$	Lively (2010b)
a_U	Density dependent effect coefficient of uninfected hosts	0.001	Lively (2010b)
a_U	Density dependent effect coefficient of infected hosts	0.001	Lively (2010b)
r_{host}	Host recombination probability	0.2	Lively (2010b)
r_{parasite}	Parasite mutation probability	0.05	Assumption
p_{host}	Probability that at least one sexual and asexual host enters the population	0.1	Assumption
p_{infected}	Probability that at least one infected host enters the population	0.02	Assumption

Table 1: **Parameter descriptions and values**. Parameters with prior distributions are estimated via Approximate Bayesian Computation (ABC). k and θ in Gamma distribution represent shape and scale parameters where mean and squared CV are given by $k\theta$ and $1/k$, respectively. α and β in Beta distribution represent shape parameters where mean and squared CV are given by $\alpha/(\alpha + \beta)$ and $\beta/(\alpha^2 + \alpha\beta + \alpha)$. N, p and θ in Beta binomial distributions represent number of trials, probability of success, and overdispersion parameters (Morris et al., 1983). We define prior on $G_{\text{asex}} - 1$ instead to always maintain at least one asexual genotype in the system. μ and σ in log-normal distribution represent mean and standard deviation on a log scale. All other parameters are fixed throughout simulations.

three tolerance values are chosen in a decreasing order to reach the final step quicker. Tolerance value of the final run is chosen so that a parameter set will be accepted if its each simulated summary statistic deviates from the corresponding observed summary statistic by 0.1 units on average. For spatial data (Dagan et al., 2013a; McKone et al., 2016), four summary statistics are compared: mean proportion of infected and sexually reproducing hosts and CV in these proportions across populations. Tolerance values of 1.6, 0.8, 0.6 and 0.4 are used for each run. For spatiotemporal data (Vergara et al., 2014b), we additionally compare the CV of proportions infected and sexually reproducing hosts across generations. Larger tolerance values (2.4, 1.2, 0.9 and 0.6) are used for each run to account for higher number of summary statistics being compared.

All statistical results are weighted by parameter weights of the final run; confidence intervals are obtained by taking weighted quantiles. However, it is always not possible to obtain the exact 95% confidence interval when the smallest (or the largest) value has a weight greater than 0.025. In these cases, we take the lowest (or largest) possible quantile that is closest to the 2.5% (or the 97.5%) quantile. While this procedure results in a slightly narrower confidence interval, the rounding errors in quantiles are negligible (for example, we use the 2.82% quantile instead of the 2.5% quantile as the lower end of a confidence interval).

2.5 Power analysis

Using estimated parameters for each data set, we calculate the power to detect a significant positive correlation between infection prevalence and frequency of sexual hosts. For each parameter sample from the final run of the ABC, 10 simulations are run. For each simulation, we take the last two generations – assuming that a year contains two snail generations (estimated generation time is 4-9 months (Neiman et al., 2005)) – from the simulation and choose n populations at random from 40 simulated populations. For each selected population, hosts are divided into four categories based on their infection status (infected/uninfected) and reproductive mode (asexual/sexual), and mean proportion of hosts in each category is calculated by averaging over two generations. Independent multinomial samples of size m are drawn from each selected population based on the proportions in each four categories. Correlation between proportion of infected hosts and proportion of sexual hosts is tested using the Spearman's rank correlation at a 5% significance level.

We caution that this procedure is essentially a post-hoc power analysis (Goodman and Berlin, 1994; Senn, 2002). While this is not as bad as a *direct* post-hoc analysis (i.e., we do not try to match observed correlations explicitly), post-hoc power analyses based on observed effect sizes are likely to overestimate power. Given that McKone et al. (2016) already observed a significant correlation, our analysis is likely to estimate high power for the same population.

3 Results

First, we compare observed summary statistics with fitted and predicted summary statistics (Fig. 1). Fitted summary statistics are those that are accepted via ABC; these can be interpreted as model-based estimates of the true summary statistics (and associated confidence intervals) of the study sites. Predicted summary statistics are obtained by simulating the model again using the estimated parameters and calculating summary statistics from randomly selected subpopulations; these values represent summary statistics that could have been obtained if other sites were chosen for the study. The predicted summary statistics are highly variable because they account for uncertainty in the unobserved subpopulations.

Our simple meta-population Red Queen model can capture observed variation in infection prevalence and frequency of sexual hosts reasonably well; both temporal and spatial variation (measured by CV across mean proportions) are well-matched by the model. However, as model fitting is performed by minimizing the sum of absoute distance between observed and simulated summary statistics, our method does not gurantee that all summary statistics are equally well-fitted. The model tends to overestimate mean proportion of infected hosts. The observed mean proportion of infected hosts are 0.175 (Dagan et al., 2013a), 0.051 (McKone et al., 2016), and 0.440 (Vergara et al., 2014b), whereas their ABC-based estimates are 0.240 (95% CI: 0.174 - 0.286), 0.313 (95% CI: 0.233 - 0.404), and 0.542 (95% CI: 0.360 - 0.730), respectively. The model underestimates mean proportion of sexual hosts for Dagan et al. (2013a) and Vergara et al. (2014b). Observed mean proportions of sexual hosts are 0.045 and 0.704, respectively, whereas their ABC-based estimates are 0.026 (95% CI: 0.007 - 0.048) and 0.596 (95% CI: 0.441 - 0.679).

To further diagnose the fit, we compare the predicted relationships between the mean proportion of infected hosts and the mean proportion of sexual hosts across subpopulations with the observed data (Fig. 2); Despite its accuracy in reproducing summary statistics reported by Dagan et al. (2013a), our model poorly captures the relationship between the mean proportion of infected hosts and the mean proportion of sexual hosts (Fig. 2; Dagan et al. (2013a)). The model predicts sexual reproduction to be well maintained when infection prevalence is high ($> 40\%$) whereas the observed data (Dagan et al., 2013a) suggests that sexual reproduction can only be supported when infection prevalence is low ($< 20\%$). McKone et al. (2016) also found sexually reproducing snails in sites with low infection prevalence ($< 20\%$).

On the other hand, Vergara et al.'s (2014b) data set suggests that sexually reproducing populations are likely to have a relatively high prevalence of infection (approximately 20% - 80%). This is broadly consistent with our model prediction. Most of the data points fall within the range of model predictions (Fig. 2; Vergara et al. (2014b)). However, there is one site in which more than 90% of the snails were found to be sexual throughout the study period of 5 years (Vergara et al., 2014b);

There is a high posterior density region in which the proportion of infected

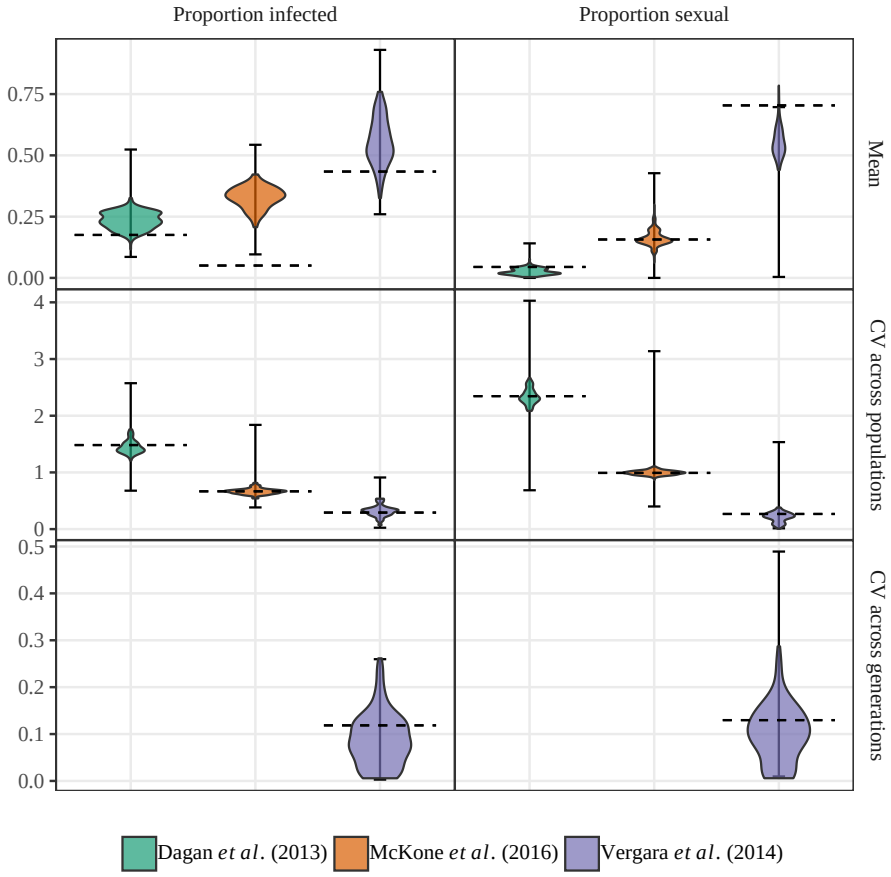

Figure 1: **Summary statistics of the observed data vs. distribution of summary statistics of the simulated data from the posterior samples.** Dotted horizontal lines represent observed summary statistics. Violin plots show weighted distribution of fitted summary statistics (i.e., summary statistics that were accepted during ABC). Error bars show 95% weighted quantiles of predicted summary statistics. The weights correspond to the posterior distribution weights from ABC. For each posterior sample, 10 simulations are run and each simulated system is sampled at random 100 times so that each sample consists of the same number of populations as in the fitted data. All univariate summary statistics are matched reasonably well, except for the mean proportion of infected hosts in McKone et al. (2016).

hosts remains almost constant (around 0.5) among subpopulations while the proportion of sexual hosts can range from 0 to 0.3 (visible in the fits to McKone et al. (2016)). As transmission rate (β) increases, selection for sexual hosts

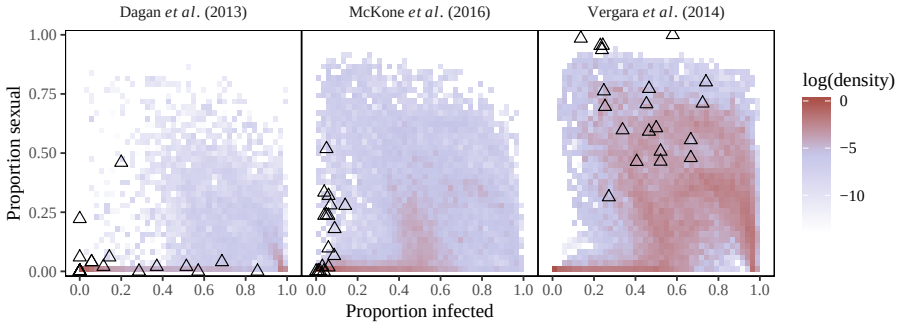

Figure 2: **Predicted relationship between mean infection prevalence and mean proportion of sexual hosts in each population.** For each posterior sample, 10 simulations are run. For each population within a simulation, mean infection prevalence and mean proportion of sexual hosts is calculated by averaging across last two generations. Each point, consisting of mean infection prevalence and meal proportion of sexual hosts, is assigned the weight of the parameter used to simulated the population. The density at each cell is calculated as the sum of weights of the points that lie within it. Cell densities are normalized by dividing by the maximum grid densities for each fit. Open triangles represent observed data; proportion of sexual hosts is computed from proportion of male hosts.

increases but the increasing number of resistant offspring prevents further infection from occuring and can decrease overall infection prevalence. This trend is consistent with previous results of Lively (2001) who noted that there is a region in which either sexual and asexual reproduction can be selected exclusively under same infection prevalence.

The proportionf of sexual hosts decreases when infection prevalence is extremely high (visible in the fits to Vergara et al. (2014b)). This pattern can be explained by the decrease in fitness of sexual hosts with increase in infection prevalence, predicted by Ashby and King (2015). The same pattern can be found in an earlier work by Lively (2010b) although it was not discussed.

Fig. 3 presents parameter estimates. Keeping in mind that we do not obtain good fits to data from Dagan et al. (2013a) and McKone et al. (2016), we still find that high virulence and a low ratio of asexual to sexual genetic diversity are necesary to explain the observed dynamics. Moreover, we are able to capture observed differences in mean and variation in infection prevalence among studies in our estimates of transmission rate parameters (β_{mean} and β_{CV}).

Our fits to McKone et al. (2016) suggest that scale parameter for the cost of sex, c_b, should be higher than our prior assumption based on Gibson et al. (2017) that the estimated cost of sex to be 2.14 (95% CI: 1.81 - 2.55). Ashby and King (2015) defined c_b as additional costs and benefits of sex, where $c_b = 1$ corresponds the two fold cost. Under their interpretation, our estimate of c_b

corresponds to a slightly lower estimate of cost of sex: 1.95 (95% CI: 1.68 - 2.4). (We propose an alternate interpretation to this parameter estimate below).

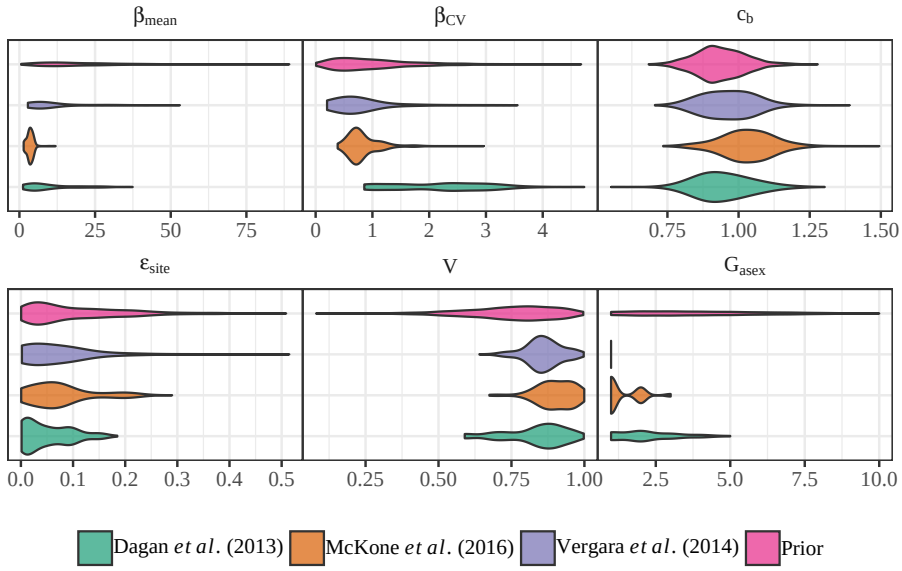

Figure 3: **Parameter estimates from Sequential Monte Carlo Approximate Bayesian Computation.** Violin plots represent weighted distribution of 100 posterior samples obtained from ABC. Violin plots for prior distribution is obtained by drawing 10000 random parameter samples from the prior distribution. G_{asex} is a discrete variable but is drawn on a continuous scale for convenience.

Finally, our power analyses suggest that there is high power to detect a positive correlation between infection prevalence and frequency of sexual hosts in the systems studied by Dagan et al. (2013a) and McKone et al. (2016) (Fig. 4). Such high power predicted for Dagan et al. (2013a) is particularly surprising given that they were not able to observe the expected correlation. This discrepancy implies that the snail populations studied by Dagan et al. (2013a) show sufficient variation in infection prevalence in order for the correlation to be observed under pure Red queen selection, but other underlying factors that are neglected by our model does not account may have caused the populations to deviate from their expected behaviours. On the other hand, our model predicts low power for detecting the positive correlation for the system studied by Vergara et al. (2014b) (Fig. 3). Overall, we predict that increasing number of sites is a more effective way to increase power than increasing number of samples per site.

While we originally planned to perform power analysis using Spearman's rank correlation, we repeated the analysis using Pearson's correlation after ap-

plying arcsine square root transformation (Lively, 1992) to see whether this procedure improves power. Using Pearson correlation gives slightly higher power to detect the positive correlation bewteen frequency of sexual hosts and infection prevalence (see Appendix).

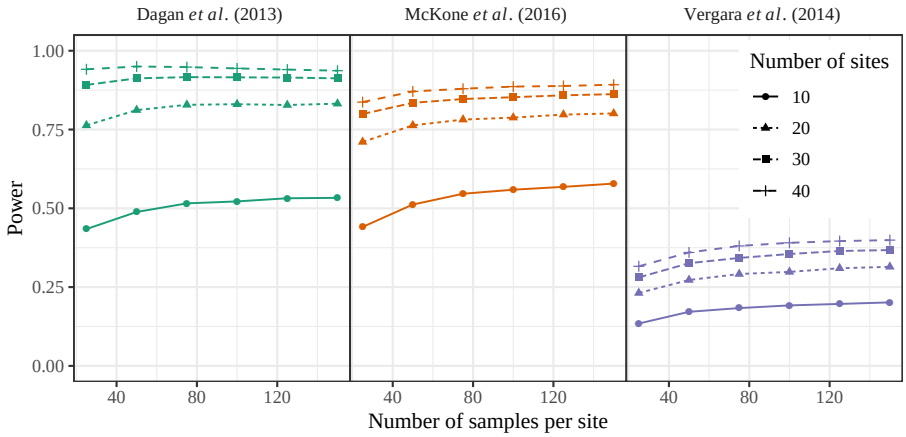

Figure 4: **Power to detect a statistically significant positive correlation between infection prevalence and frequency of sexual hosts.** Spearman's rank correlation was used to test for correlation between infection prevalence and frequency of sexual hosts in simulated data from the posterior distributions.

4 Discussion

A simple metapopulation model for host-parasite coevolution is able to match the observed prevalence of sexual reproduction and trematode infection in snail populations (Fig. 1). A direct comparison between a model and a data allows us to infer biologically meaningful parameters of a model (Fig. 3) and estimate the power to observe a significant, positive correlation between the proportion of infected hosts and the proportion of sexual hosts (Fig. 4), as predicted by Lively (1992). However, discrepancies between the model predictions and observed data suggest that a simple host-parasite coevolution model may not be able to sufficiently explain maintenance of sexual reproduction observed in snail populations (Fig. 2).

A model that fits poorly can sometimes tell us more about a biological system than a model that fits well. Our model clearly failed to match the data presented by Dagan et al. (2013a) (Fig. 2). The snail populations studied by Dagan et al. (2013a) live in qualitatively different environments from two other snail populations that we considered (Vergara et al., 2014b; McKone et al., 2016). For example, their habitats are subject to seasonal flash floods (Ben-Ami

and Heller, 2007), which can affect reproductive strategies of snails and interfere with the host-parasite coevolution (Lytle, 2000; Ben-Ami and Heller, 2007). As a result, a positive correlation between infection prevalence and frequency of sexual reproduction could not be detected from the system even though our model (which does not incorporate disturbance) predicts high power. These results suggest that a significant correlation (or the lack of it) between infection prevalence and the frequency of sexual reproduction may not provide sufficient evidence for (or against) the role of parasites in maintaining sexual reproduction in the host population.

The model fits to Dagan et al. (2013a) and McKone et al. (2016) suggests that cost of sex can be overcome and sexual reproduction can be maintained only if infection prevalence is much higher than the observed prevalence (Fig. 2). In other words, benefit of producing offspring with novel genotype is relatively small when infection prevalence is low. In order to support sexual reproduction at lower infection prevalence, the benefit of sex must be greater or other mechanisms must compensate for the difference. As our model relies on a simple structure and strong parametric assumptions, additional benefit of sex can only be provided by lowering the cost of sex (i.e., increasing the scale parameter, c_b).

The simple structure of the model and limited genetic diversity can explain the discrepancy observed in model prediction and the observed data by McKone et al. (2016). Here, we assumed that host resistance to infection is deteremined entirely by two biallelic loci, resulting in 10 possible genotypes. However, it is unlikely that such simple model can capture the genetic interaction between hosts and parasites observed in nature. Although exact genetic architecture that determines trematode infecion in snails (e.g., the number and inheritence of loci involved in parasite resistance) is not known, the documented genetic diversity of snails is far greater than what we have assumed (Fox et al., 1996; King et al., 2011; Dagan et al., 2013a). Increasing the maximum possible genetic diversity of the model would have allowed sexual hosts to escape infection more easily and maintained sexual reproduction at lower prevalence of infection (Lively, 2010a; King and Lively, 2012; Ashby and King, 2015).

While the positive correlation between frequency of sexual hosts and prevalence of infection can provide evidence for the effect of the parasite on the maintenance of sexual reproduction in the host population, it does not fully capture the coevolutionary process. In particular, the positive correlation represents a contrast between populations that undergo Red Queen dynamics and those that do not (rather than a continuous relationship between the frequency of sexual hosts and prevalence of infection): low-risk populations will be dominated by asexual individuals whereas high-risk populations will consist of both sexual and asexual individuals. In giving a similar explanation, Lively (2001) noted that large variation in the risk of infection is required to observe the predicted correlation.

The similarity between the correlations predicted by the model and the correlations observed in nature may be interpreted as providing more confidence that the observed popoulations provide evidence for the Red Queen Hypothesis (Fig. 5). However, the strength of correlation does not provide any evidence

for the strength of selection imposed on the host population. Moreover, when populations are going through active coevolution, wide range of correlations can be detected (Fig. 5: Vergara et al. (2014b)). While simple statistical tools, such as a correlation coefficient, are easily accessible, more sophisticated and mechanistic statistical models that yield biologically interpretable parameters may be better suited for understanding the effect of parasites on the maintenance of sexual reproduction in the host population.

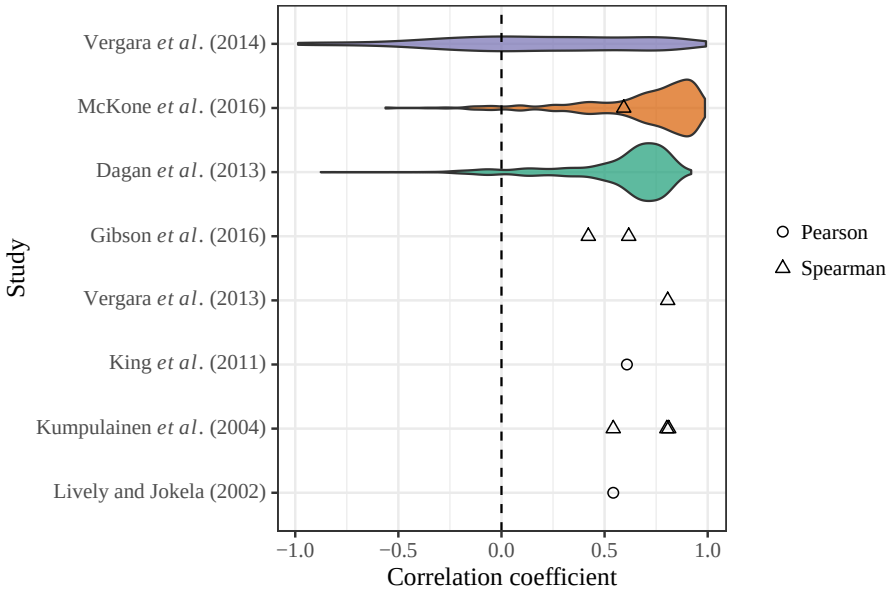

Figure 5: **Predicted vs. observed correlation between infection prevalence and frequency of sexual hosts.** Violin plots show weighted distribution of predicted strenght of correlation. Predicted vorrelation was measured for each simulation from the posterior by taking into account last two generations from each simulated population. Triangles and circles represent observed correlation sizes from previous studies that found positive correlation.

Mathematical models have used extensively to build theoretical foundations for the evolution of sex but only a few models have been confronted with data. The idea of two fold cost of sex was only recently tested directly by fitting a theoretical model to observed data (Gibson et al., 2017). Making statistical inference on the observed systems and testing theoretical models against data may provide deeper insight into underpinnings of maintenance of sexual reproduction.

5 Acknowledgements

We thank SHARCnet for providing computational resources.

6 Funding

This work is supported by The Natural Sciences and Engineering Research Council, Undergraduate Student Research Award (to SWP).

References

Agrawal, A. F. and C. M. Lively (2001). Parasites and the evolution of self-fertilization. *Evolution 55*(5), 869–879.

Ashby, B. and S. Gupta (2014). Parasitic castration promotes coevolutionary cycling but also imposes a cost on sex. *Evolution 68*(8), 2234–2244.

Ashby, B., R. Iritani, A. Best, A. White, and M. Boots (2019). Understanding the role of eco-evolutionary feedbacks in host-parasite coevolution. *Journal of theoretical biology 464*, 115–125.

Ashby, B. and K. C. King (2015). Diversity and the maintenance of sex by parasites. *Journal of Evolutionary Biology 28*(3), 511–520.

Auld, S. K., S. K. Tinkler, and M. C. Tinsley (2016). Sex as a strategy against rapidly evolving parasites. *Proceedings of the Royal Society B: Biological Sciences 283*(1845).

Bell, G. (1982). *The Masterpiece of Nature: The Evolution and Genetics of Sexuality*. University of California Press.

Ben-Ami, F. (2006). First report of the invasive freshwater snail *Tarebia granifera* (Lamarck, 1816) (Gastropoda: Thiaridae) from Israel. *Nautilus 120*(4), 156–161.

Ben-Ami, F., D. Gold, and B. Fried (2005). Differential infectivity of *Transversotrema patialense* for naive fish. *Journal of Parasitology 91*(4), 949–951.

Ben-Ami, F. and J. Heller (2005). Spatial and temporal patterns of parthenogenesis and parasitism in the freshwater snail *Melanoides tuberculata*. *Journal of Evolutionary Biology 18*(1), 138–146.

Ben-Ami, F. and J. Heller (2007). Temporal patterns of geographic parthenogenesis in a freshwater snail. *Biological journal of the Linnean Society 91*(4), 711–718.

Ben-Ami, F. and J. Heller (2008). Sex versus parasitism versus density. *Biological journal of the Linnean Society 93*(3), 537–544.

Buckling, A. and P. B. Rainey (2002). Antagonistic coevolution between a bacterium and a bacteriophage. *Proceedings of the Royal Society of London. Series B: Biological Sciences 269*(1494), 931–936.

Clarke, B. (1976). The ecological genetics of host-parasite relationships. *Genetic aspects of host-parasite relationships. Blackwell, London*, 87–103.

Dagan, Y., K. Liljeroos, J. Jokela, and F. Ben-Ami (2013a). Clonal diversity driven by parasitism in a freshwater snail. *Journal of evolutionary biology 26*(11), 2509–2519.

Dagan, Y., K. Liljeroos, J. Jokela, and F. Ben-Ami (2013b). Data from: Clonal diversity driven by parasitism in a freshwater snail.

Decaestecker, E., S. Gaba, J. A. Raeymaekers, R. Stoks, L. Van Kerckhoven, D. Ebert, and L. De Meester (2007). Host–parasite 'Red Queen'dynamics archived in pond sediment. *Nature 450*(7171), 870.

Dybdahl, M. F. and C. M. Lively (1995a). Diverse, endemic and polyphyletic clones in mixed populations of a freshwater snail (*Potamopyrgus antipodarum*). *Journal of Evolutionary Biology 8*(3), 385–398.

Dybdahl, M. F. and C. M. Lively (1995b). Host-parasite interactions: infection of common clones in natural populations of a freshwater snail (*Potamopyrgus antipodarum*). *Proceedings of the Royal Society of London B: Biological Sciences 260*(1357), 99–103.

Dybdahl, M. F. and C. M. Lively (1998). Host-parasite coevolution: evidence for rare advantage and time-lagged selection in a natural population. *Evolution*, 1057–1066.

Elderd, B. D., V. M. Dukic, and G. Dwyer (2006). Uncertainty in predictions of disease spread and public health responses to bioterrorism and emerging diseases. *Proceedings of the National Academy of Sciences 103*(42), 15693–15697.

Fox, J. A., M. F. Dybdahl, J. Jokela, and C. M. Lively (1996). Genetic structure of coexisting sexual and clonal subpopulations in a freshwater snail (*Potamopyrgus antipodarum*). *Evolution 50*(4), 1541–1548.

Gibson, A. K., L. F. Delph, and C. M. Lively (2017). The two-fold cost of sex: Experimental evidence from a natural system. *Evolution Letters 1*(1), 6–15.

Gibson, A. K., L. F. Delph, D. Vergara, and C. M. Lively (2018). Periodic, parasite-mediated selection for and against sex. *The American Naturalist 192*(5), 537–551.

Gibson, A. K., J. Y. Xu, and C. M. Lively (2016). Within-population covariation between sexual reproduction and susceptibility to local parasites. *Evolution 70*(9), 2049–2060.

Goodman, S. N. and J. A. Berlin (1994). The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. *Annals of internal medicine 121*(3), 200–206.

Haafke, J., M. Abou Chakra, and L. Becks (2016). Eco-evolutionary feedback promotes red queen dynamics and selects for sex in predator populations. *Evolution 70*(3), 641–652.

Haldane, J. B. S. (1949). Disease and evolution. *La Ricerca Scientific Supplement 19*, 68–76.

Hamilton, W. D. (1980). Sex versus non-sex versus parasite. *Oikos*, 282–290.

Hamilton, W. D., R. Axelrod, and R. Tanese (1990). Sexual reproduction as an adaptation to resist parasites (a review). *Proceedings of the National Academy of Sciences 87*(9), 3566–3573.

Heller, J. and V. Farstey (1990). Sexual and parthenogenetic populations of the freshwater snail melanoides tuberculata in Israel. *Israel Journal of Ecology and Evolution 37*(2), 75–87.

Howard, R. S., C. M. Lively, et al. (1994). Parasitism, mutation accumulation and the maintenance of sex. *Nature 367*(6463), 554–557.

Jaenike, J. (1978). An hypothesis to account for the maintenance of sex within populations. *Evolutionary Theory 3*, 191–194.

Jokela, J., M. F. Dybdahl, and C. M. Lively (2009). The maintenance of sex, clonal dynamics, and host-parasite coevolution in a mixed population of sexual and asexual snails. *The American Naturalist 174*(S1), S43–S53.

King, K. and C. M. Lively (2012). Does genetic diversity limit disease spread in natural host populations? *Heredity 109*(4), 199–203.

King, K. C., L. F. Delph, J. Jokela, and C. M. Lively (2011). Coevolutionary hotspots and coldspots for host sex and parasite local adaptation in a snail-trematode interaction. *Oikos 120*(9), 1335–1340.

King, K. C., J. Jokela, and C. M. Lively (2011). Parasites, sex, and clonal diversity in natural snail populations. *Evolution 65*(5), 1474–1481.

Koskella, B. (2013). Phage-mediated selection on microbiota of a long-lived host. *Current Biology 23*(13), 1256–1260.

Koskella, B. (2014). Bacteria-phage interactions across time and space: merging local adaptation and time-shift experiments to understand phage evolution. *The American Naturalist 184*(S1), S9–S21.

Koskella, B. and C. M. Lively (2009). Evidence for negative frequency-dependent selection during experimental coevolution of a freshwater snail and a sterilizing trematode. *Evolution: International Journal of Organic Evolution 63*(9), 2213–2221.

Kumpulainen, T., A. Grapputo, J. Mappes, and M. Björklund (2004). Parasites and sexual reproduction in psychid moths. *Evolution 58*(7), 1511–1520.

Lehtonen, J., M. D. Jennions, and H. Kokko (2012). The many costs of sex. *Trends in Ecology & Evolution 27*(3), 172–178.

Lively, C. (2009). The maintenance of sex: host–parasite coevolution with density-dependent virulence. *Journal of Evolutionary Biology 22*(10), 2086–2093.

Lively, C. M. (1987). Evidence from a New Zealand snail for the maintenance of sex by parasitism. *Nature 328*(6130), 519–521.

Lively, C. M. (1989). Adaptation by a parasitic trematode to local populations of its snail host. *Evolution 43*(8), 1663–1671.

Lively, C. M. (1992). Parthenogenesis in a freshwater snail: reproductive assurance versus parasitic release. *Evolution 46*(4), 907–913.

Lively, C. M. (2001). Trematode infection and the distribution and dynamics of parthenogenetic snail populations. *Parasitology 123*(07), 19–26.

Lively, C. M. (2010a). The effect of host genetic diversity on disease spread. *The American Naturalist 175*(6), E149–E152.

Lively, C. M. (2010b). An epidemiological model of host–parasite coevolution and sex. *Journal of evolutionary biology 23*(7), 1490–1497.

Lively, C. M. (2010c). A review of Red Queen models for the persistence of obligate sexual reproduction. *Journal of Heredity 101*(suppl_1), S13–S20.

Lively, C. M. (2018). Habitat heterogeneity, host population structure, and parasite local adaptation. *Journal of Heredity 109*(1), 29–37.

Lively, C. M., M. F. Dybdahl, J. Jokela, E. E. Osnas, and L. F. Delph (2004). Host sex and local adaptation by parasites in a snail-trematode interaction. *the american naturalist 164*(S5), S6–S18.

Lively, C. M. and R. S. Howard (1994). Selection by parasites for clonal diversity and mixed mating. In *Infection, Polymorphism and Evolution*, pp. 1–11. Springer.

Lively, C. M. and J. Jokela (2002). Temporal and spatial distributions of parasites and sex in a freshwater snail. *Evolutionary Ecology Research 4*(2), 219–226.

Lytle, D. A. (2000). Biotic and abiotic effects of flash flooding in a montane desert stream. *Archiv für Hydrobiologie*, 85–100.

MacPherson, A. and S. P. Otto (2018). Joint coevolutionary–epidemiological models dampen red queen cycles and alter conditions for epidemics. *Theoretical population biology 122*, 137–148.

May, R. M. and R. M. Anderson (1983). Epidemiology and genetics in the coevolution of parasites and hosts. *Proceedings of the Royal Society of London B: Biological Sciences 219*(1216), 281–313.

McArthur, C. P. and D. Featherston (1976). Suppression of egg production in Potamopyrgus antipodarum (Gastropoda: Hydrobiidae) by larval trematodes. *New Zealand Journal of Zoology 3*(1), 35–38.

McKone, M. J., A. K. Gibson, D. Cook, L. A. Freymiller, D. Mishkind, A. Quinlan, J. M. York, C. M. Lively, and M. Neiman (2016). Fine-scale association between parasites and sex in *Potamopyrgus antipodarum* within a New Zealand lake. *New Zealand Journal of Ecology 40*(3), 1.

Morran, L. T., R. C. Parrish, I. A. Gelarden, M. B. Allen, and C. M. Lively (2014). Experimental coevolution: rapid local adaptation by parasites depends on host mating system. *The American naturalist 184*(S1), S91–S100.

Morris, C. N. et al. (1983). Natural exponential families with quadratic variance functions: statistical theory. *The Annals of Statistics 11*(2), 515–529.

Neiman, M., J. Jokela, and C. Lively (2005). Variation in asexual lineage age in *Potamopyrgus antipodarum*, a New Zealand snail. *Evolution 59*(9), 1945–1952.

Otto, S. P. and Y. Michalakis (1998). The evolution of recombination in changing environments. *Trends in Ecology & Evolution 13*(4), 145–151.

Phillips, N. R. and D. M. Lambert (1989). Genetics of *Potamopyrgus antipodarum* (Gastropoda: Prosobranchia): evidence for reproductive modes. *New Zealand journal of zoology 16*(3), 435–445.

Rauch, G., M. Kalbe, and T. B. Reusch (2006). One day is enough: rapid and specific host–parasite interactions in a stickleback-trematode system. *Biology Letters 2*(3), 382–384.

Samadi, S., J. Mavárez, J.-P. Pointier, B. Delay, and P. Jarne (1999). Microsatellite and morphological analysis of population structure in the parthenogenetic freshwater snail *Melanoides tuberculata*: insights into the creation of clonal variability. *Molecular Ecology 8*(7), 1141–1153.

Senn, S. J. (2002). Power is indeed irrelevant in interpreting completed studies. *Bmj 325*(7375), 1304.

Slowinski, S. P., L. T. Morran, R. C. Parrish, E. R. Cui, A. Bhattacharya, C. M. Lively, and P. C. Phillips (2016). Coevolutionary interactions with parasites constrain the spread of self-fertilization into outcrossing host populations. *Evolution 70*(11), 2632–2639.

Smith, J. M. (1978). *The Evolution of Sex*, Volume 54. Cambridge Univ Press.

Smith, J. M. and M. Slatkin (1973). The stability of predator-prey systems. *Ecology 54*(2), 384–391.

Song, Y., C. S. Gokhale, A. Papkou, H. Schulenburg, and A. Traulsen (2015). Host-parasite coevolution in populations of constant and variable size. *BMC evolutionary biology 15*(1), 212.

Thrall, P. H., A.-L. Laine, M. Ravensdale, A. Nemri, P. N. Dodds, L. G. Barrett, and J. J. Burdon (2012). Rapid genetic change underpins antagonistic coevolution in a natural host-pathogen metapopulation. *Ecology Letters 15*(5), 425–435.

Toni, T., D. Welch, N. Strelkowa, A. Ipsen, and M. P. Stumpf (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface 6*(31), 187–202.

Turner, B. M. and T. Van Zandt (2012). A tutorial on approximate bayesian computation. *Journal of Mathematical Psychology 56*(2), 69–85.

Vergara, D., J. Jokela, and C. Lively (2014a). Data from: Infection dynamics in coexisting sexual and asexual host populations: support for the Red Queen hypothesis.

Vergara, D., J. Jokela, and C. M. Lively (2014b). Infection dynamics in coexisting sexual and asexual host populations: support for the Red Queen hypothesis. *The American naturalist 184*(S1), S22–S30.

Vergara, D., C. M. Lively, K. C. King, and J. Jokela (2013). The geographic mosaic of sex and infection in lake populations of a New Zealand snail at multiple spatial scales. *The American Naturalist 182*(4), 484–493.

Verhoeven, K. J. and A. Biere (2013). Geographic parthenogenesis and plant-enemy interactions in the common dandelion. *BMC evolutionary biology 13*(1), 23.

Vrijenhoek, R. C. (1998). Animal clones and diversity. *Bioscience 48*(8), 617–628.

Wallace, C. (1992). Parthenogenesis, sex and chromosomes in *Potamopyrgus*. *Journal of Molluscan Studies 58*(2), 93–107.

Winterbourn, M. (1974). Larval Trematoda parasitizing the New Zealand species of *Potamopyrgus* (Gastropoda: Hydrobiidae). *Mauri Ora 2*, 17–30.

Appendix

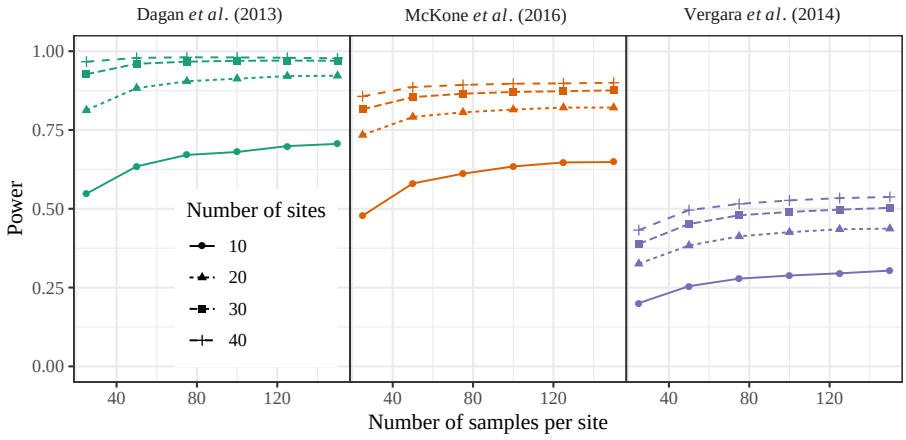

Figure A1: **Power to detect a statistically significant positive correlation between infection prevalence and frequency of sexual hosts.** Pearson correlation was used to test for correlation between square root arcsine transformed infection prevalence and frequency of sexual hosts in simulated data from the posterior distributions.