Needs a title

Sang Woo Park, Benjamin M Bolker

June 22, 2018

1 Introduction

The evolution of sexual reproduction still remains to be explained (Otto, 2009). Despite being the dominant mode of reproduction (Vrijenhoek, 1998), sexual reproduction entails numerous costs (Lehtonen et al., 2012). Perhaps the most often mentioned is the cost of producing males (Smith, 1978). As males cannot produce offsprings, sexual population is expected to be outgrown by the asexual counterpart that can grow twice as fast (assuming that the sexual population produces 50% male and 50% female). The infamous *two-fold cost of sex* (Smith, 1978) relies on the assumption that everything else is equal but this is not necessarily true. So what drives sex to persist?

One explanation for persistence of sexual reproduction is the Red Queen Hypothesis (Bell, 1982). The Red Queen Hypothesis predicts sexually reproducing hosts to overcome the cost of sex under strong parasite selection by producing genetically diverse offspring that are resistant to infection (Haldane, 1949; Jaenike, 1978; Hamilton, 1980). Host parasite coevolution constantly creates selective advantage for rare genotypes, hence creating osccilation in genotypic frequencies, and allows for sexual reproduction to persist in the host population (Clarke, 1976; Hamilton, 1980).

Since the conception of the Red Queen Hypothesis, much of the theoretical work has focused on figuring out conditions under which parasite selection can maintain sexual reproduction in the host population. May and Anderson (1983) first noted that parasites must be extremely virulent to maintain sexual reproduction but later studies showed that sexual and asexual hosts can coexist even at intermediate virulence (Howard et al., 1994). Agrawal and Lively (2002) compared wide range of infection genetics that determine parasite resistance and the dynamics that arise from different genetic architecture. Recent studies showed that host genetic diversity plays an important role in determining the selection for sex [CITE].

Some theoretical studies took a departure from the classical population genetics framework and studied effects of ecological and epidemiological on the Red Queen dynamics. On one hand, number of studies showed that incorporating ecological and epidemiological details can support sexual reproduction in the host population (**?**Galvani et al., 2003; Lively, 2009, 2010b). On the other

hand, MacPherson and Otto (2017) showed that Red Queen dynamics (i.e., cycles in allele frequencies) fails to persist when explicit epidemiological structure is taken into account with coevolutionary dynamics.

Empiricists have mostly focused on confirming predictions that stem from the Red Queen hypothesis. Typical among them are local adaptation, time-lagged selection, and association between parasites and host reproductive mode (see Tobler and Schlupp (2008) and Vergara et al. (2014b) for reviews). An important example is the snail population in New Zealand that serve as an intermediate hosts for trematodes [CITE]. Through several decades of work, Lively *et al.*demonstrated that the population satisfies necessary conditions for the host-parasite coevolutionary dynamics and provide support for the hypothesis (Lively, 1987, 1989; Dybdahl and Lively, 1995, 1998; Jokela et al., 2009; Vergara et al., 2014b; Gibson et al., 2016). Recently, Auld et al. (2016) and **?** showed that more direct evidence can be achieved using experimental systems.

While the Red Queen Hypothesis has gained reasponable support both theoretically and empirically, there still remains a gap between theory and data. Many models for Red Queen Hypothesis rely on simplifying assumption that are not applicable to natural populations and make predictions based on assumed parameters. In particular, none of the Red Queen models reviewed by Ashby and King (2015) use statistical tools to relate the model to data. However, there are exceptions: Lively (1992) initially postulated that infection prevalence should be positively correlated with frequency of sexual hosts. (Lively, 2001) then formalized the idea with a mathematical model and confirmed the prediction with a snail-trematode system. Many other emprical works have successfully found the positive correlation (Lively and Jokela, 2002; Kumpulainen et al., 2004; Vergara et al., 2013; McKone et al., 2016; Gibson et al., 2016) but surprisingly such correlation was not observed in a different snail-trematode system (Dagan et al., 2013a).

Here, we try to bridge the gap between theory and data further. We extend the model used by Lively (2010b) to include stochasticity and population structure. Then, we fit it the to data from Dagan et al. (2013a); McKone et al. (2016); Vergara et al. (2014b) using Approximate Bayesian Computation (ABC) to estimate biologically relavent parameters. Using estimated parameters obtained from the fits, we assess the model by comparing its fit to data and perform a power analysis to test the prediction that infection prevalence is positively correlated with frequency of sexual reproduction Lively (2001).

2 Methods

2.1 Study systems

We consider two snail-trematode populations in New Zealand (Vergara et al., 2014b; McKone et al., 2016) and a similar snail-trematode population in Isral (Dagan et al., 2013a). The snail-trematode system has been extensively studied under the context of the Red Queen Hypothesis [CITE] so we expect a simple

Red Queen model to fit well. Data collected by Dagan et al. (2013a) and Vergara et al. (2014b) were obtained from their Dryad repositories (Dagan et al., 2013b; Vergara et al., 2014a) and data by McKone et al. (2016) was extracted from their figure.

2.2 Model

We model obligate sexual hosts competing with obligate asexual hosts in multiple population by extending the model introduced by Lively (2010b). A model that Lively (2010b) used is essentially an SI model on a discrete time scale with natural mortality and virulence (reduced offspring production by infected hosts). It is a suitable canditate for this study as it captures essential structures that are present in basic epidemiological and population biology models and is general enough to be applied to broad range of natural systems. We do not consider mechanistic details of the snail-trematode system such as life history of trematodes (Vergara et al., 2014b). We incorporate population structure and allow for mixing between populations. Each population can be considered as a sampling site in the observed populations.

All hosts are assumed to be diploids with two biallelic loci, 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. For simplicity, we drop the superscript representing population and write $S_{ij}(t)$ and $A_{ij}(t)$; every population is modeled using the same equations unless noted otherwise. Following Lively (2010b), the expected number of offsprings (without recombination or outcrossing) produced by a sexual population is given by

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

where s is the proportion of males produced by sexual hosts, $S_{ij,U}$ 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. We allow for cost of sex to vary by multiplying a scale parameter, c_b, to the growth term ($2/c_b$ corresponds to two fold cost of sex) (Ashby and King, 2015). Equivalently, S'_{ij} is the amount of genotypic contribution by sexual hosts to the next generation.

Asexual hosts are assumed to be strictly clonal. Then, the expected number of offsprings produced by an asexual population 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 mixing between populations occur at rate ϵ_{mix}. Then, the expected number of offspring in the next generation (accounting for contributions

from all populations) is given by

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

where $f_{\mathrm{sex}}(x)$ is the function that models sexual reproduction, including recombination at a rate r_{host} and outcrossing, and $n_{\mathrm{population}}$ is the number of populations modeled. We then take a poisson random variable to simulate process error and allow for stochastic migration 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 probability of a sexual and an asexual host with genotype ij entering the popualation.

Infection is modeled using the matching alleles model (Otto and Michalakis, 1998). We assume that heterozygous individuals are equally susceptible to parasites that match either haplotypes. The expected number of infected hosts that carry parasite with genotype i at generation t is given by:

$$I_i(t) = \sum_{p} 2^{\delta_{ij}}\left(S_{ip,i,I}(t) + A_{ip,i,I}(t)\right), \tag{5}$$

where δ_{ij} is kronecker delta. $S_{ip,i,I}$ and $A_{ip,i,I}$ represent the expected numbers of sexual and asexual hosts infected with genotype i parasite. Following Ashby and King (2015), we assume that mutation can occur in one locus with probability r_{parasite}. We also allow for external migration of an infected host carrying parasite i with probability $p_{i,\mathrm{parasite}}$ to avoid fixation.

The amount of infectious contacts made by infected hosts from a population is given by $\lambda_i^k = \beta^k I_i'^k(t)$, where β^K is the transmission rate at each population, and $I_i'(t)$ is the number of infected hosts accounting for mutation and migration. Since we allow for mixing between population, infected hosts in a population can make contact to susceptible hosts in other populations as well. Then, all infectious contacts made by infected hosts that applies to a host in a single population is

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

Then, the force of infection that a susceptible host experiences is given by

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

The probability of infection for a host with genotype ij at site k in the next generation is

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

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

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

Expected number of individuals infected by parasites with genotype i in the next generation is given by a ratio of λ:

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

2.3 Simulation design and parameterization

Most Red Queen models have studied a competition between a single asexual genotype and sexual genotypes or have assumed equal genetic diversity of asexual and sexual hosts (Ashby and King, 2015) but neither of these assumptions are realistic. Instead, Ashby and King (2015) adopted a more realistic approach by allowing for stochastic migration of an asexual genotype to a population. Here, we combine these methods. We allow for stochastic migration of asexual hosts with different genotypes but fix number of asexual genotypes (denoted by G_{asex}) that can be present in the system, while the number of sexual genotypes (G_{sex}) that can be present in the population remains equal to the size of the genotypic space (= 10 in diploid hosts with two biallelic loci).

In the beginning of every simlation, G_{asex} is randomly chosen between 1 and 10 and asexual genotypes that can be introduced to the population are uniformly sampled from the entire genotypic space. 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 and asexual host enters the population in a generation. We scale the probability of infected host carrying parsite genotype i in a similar way for interpretability:

$$p_{i,\text{parasite}} = 1 - (1 - p_{\text{infected}})^{1/4}, \tag{12}$$

where p_{infected} is the probability that at least one infected host enters the population in a generation.

Each simulation consists of 40 populations. Every population consists of 2000 sexual hosts where 80 of them are infected. They are assumed to be in Hardy-Weinberg equilibrium with ratio between alleles being exactly half.

Transmission rate parameter, β, is randomly drawn for each population 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 (note that asexual genotype introduced can vary across population) and simulation runs for 600 generations while allowing for stochastic migration of asexuals as well.

2.4 Approximate Bayesian Computation

We use Approximate Bayesian Computation (ABC) to fit the model (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 and time as summary statistics of interest. CV of these proportions across space is calculated by first calculating mean proportions by averaging across generations and then taking the the CV of these mean proportions. Likewise, CV across generation is calculated by first averaging across space and then taking the CV.

We use uninformative priors for all parameters except c_b, a scale parameter for the cost of sex (See Table [TODO]). The prior distribution for the scale parameter is chosen so that 95% quantile of cost of sex ($2/c_b$) is approximately equal to the 95% confidence interval reported by Gibson et al. (2017). As Dagan et al. (2013a) and McKone et al. (2016) report proportion of males, instead of proportion of sexual hosts, we double the proportion of males reported when we fit the model.

To allow for more efficient estimation of the posterior distribution, we use the Population Monte Carlo approach Turner and Van Zandt (2012). First, we start by performing basic ABC. For each random parameter sample drawn from the prior distribution, the model is simulated and a sample of simulated populations is drawn from the simulated system such that the number of sampled population is equal to the number of sites collected in a study. Then, summary statistics are calculated based on the last 100 generations and the parameter is accepted if the distance between simulated and observed data (measureed by the sum of absolute differences in summary statistics) is less than the tolerance value. This process is repeated until 100 parameters are accepted.

After the first run ($t = 1$), equal weights ($w_{i,1} = 1/100$) are assigned to each accepted parameter $\theta_{i,1}$. For any run $t > 1$, a weighted random sample (θ^*) is drawn from the accepted parameters of the previous run ($t-1$) and a parameter sample ($\theta_{i,t}$) is proposed from a multivariate normal distribution with a mean θ^* and a variance covariance matrix that is equal to $\sigma^2_{t-1} = 2\text{Var}(\theta_{1:N,t-1})$, where $\text{Var}(\theta_{1:N,t-1})$ is the weighted variance covariance matrix. G_{asex} is rounded to the nearest integer and the model is simulated. If a proposed parameter is

accepted, the following weight is assigned:

$$w_{i,t} = \frac{\pi(\theta_{i,t})}{\sum_{i=1}^{100} w_{j,t-1} q(\theta_{j,t-1}|\theta_{i,t}, \sigma_{t-1}^2)}$$

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

For each data, we perform 4 runs with decreasing tolerance every run. For spatial data (Dagan et al., 2013a; McKone et al., 2016), tolerances of 1.6, 0.8, 0.6 and 0.4 are used for each run. For spatiotemporal data (Vergara et al., 2014b), tolerance of 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.

2.5 Power analysis

To calculate the power to detect positive correlation between infection prevalence and frequency of sexual hosts, we simulate an observed sample by drawing a random multinomial sample from simulated data. For simplicity, we assume that a year contains of two snail generations [CITE] and that samples are taken within a short period of time. For each simulation from the posterior distributions, we start by setting a reference generation to generatation 1001 and choosing n populations at random from 40 simulated popoulations. 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 consecutive generations. A multinomial sample of size m is drawn from each selected population independently based on the four categories. Positive correlation between proportion of infected hosts and proportion of sexual hosts is tested using Spearman's rank correlation at 5% significance level. We repeat the process 100 times for each simulation from a posterior distribution.

3 Results

Our simple Red Queen model with population structure can capture observed variation in infection prevalence and frequency of sexual hosts reasonably well. Observed coefficients of variation in mean infection prevalence and mean frequency of sexual hosts are well-matched by the same summary statistics from posterior simulations (Fig. 1). As we account for uncertainty in unboserved sites by simulating greater number of populations, there is large variation in simulated summary statistics. Dagan et al. (2013a), McKone et al. (2016) and Vergara et al. (2014b) reported mean infection prevalence of 17.5%, 5.1% and 44%, respectively, but the corresponding mean (95% quantile) estimate of the fitted model are 29.1% (8.1% - 55.7%), 45.2% (17.3% - 73.9%) and 68% (26.1% - 95.3%), respectively. The model particularly underestimates mean frequency

of sexual hosts (mean (95% quantile): 32.2% (25.3% - 64.7%)) when fitted to Vergara et al. (2014b) population (70.3%). Even though model fitting was performed by minimizing the absolute distance between observed and simulated summary statistics, there is large variation in summary statistics of the posterior simulations because uncertainty in unobserved populations is taken into account.

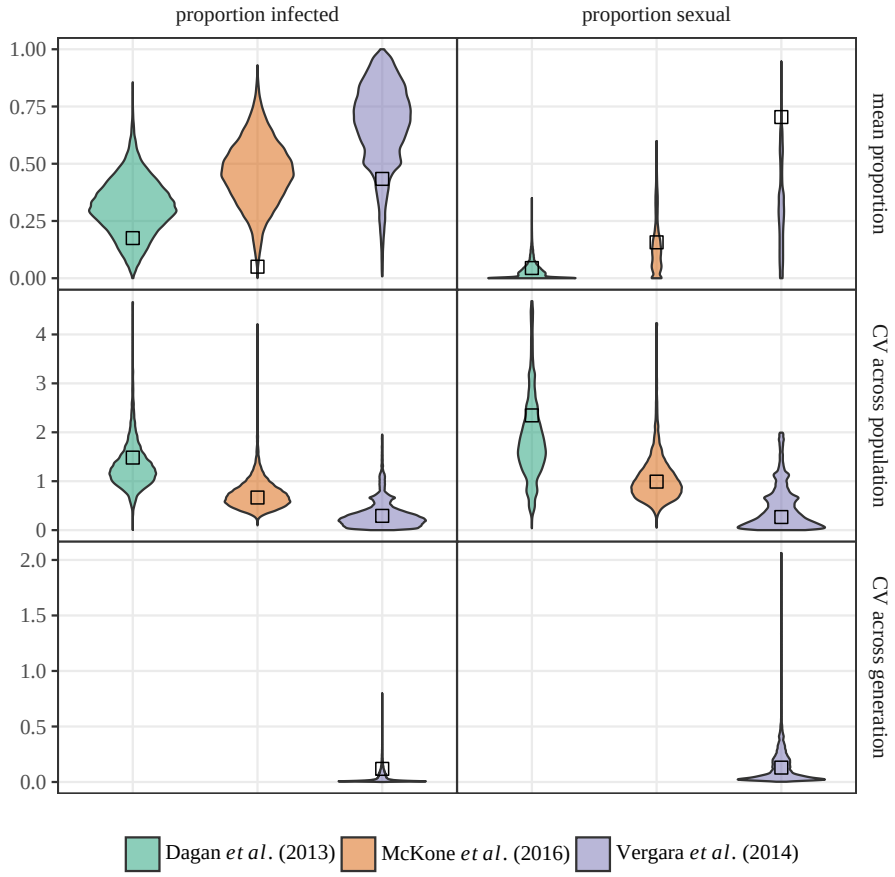

Figure 1: **Summary statistics of the observed data v.s. distribution of summary statistics of the simulated data from the posterior samples.** Violin plots show weighted distribution of summary statistics from simulated data. For each posterior sample, 10 simulations are ran and each simulated system is sampled at random 100 times so that each sample consists of equal number of populations as number of sites in fitted data. Then, summary statistics are calculated for each sample and are weighted by particle weights. Squares represnt observed summary statistics.

To further diagnose the fit, we compare the relationships between infection prevalence and frequency of sexual hosts predicted by the model with the observed data (Fig. 2). Despite being able to reproduce summary statistics reported by Dagan et al. (2013a) well, our model is unable to predict qualitative trend bewteen proportion of sexual hosts and proportion of infected hosts (Fig. 2; Dagan et al. (2013a)). While both simulated data and observed data mostly consist of asexual populations, our model predicts sexual reproduction to be maintained when infection prevalence is relatively high. On the other hand, Dagan et al. (2013a) data suggests that sexual reproduction is only maintained when infection prevalence much lowere. Similarly, overestimation of infection prevalence is strongly pronounced in our fits to McKone et al. (2016).

While there are a few data points that appear to be outliers compared to our fits to Vergara et al. (2014b), it is important to note that Fig. 2 only shows proportion of infected and sexual snails averaged across 100 generations for each site. The observed data are more likely to be samples across a few generations and the cyclic nature of the Red Queen dynamics creates more variation in the data, which is not captured in Fig. 2. On the other hand, Vergara et al. (2014b) reported greater than 90% sexual snails throughout 5 years in one site but our model can only maintain up to 75% sexual snails on average given similar inection prevalence. Based on the goodness of fits, we exclude fits to Dagan et al. (2013a) and McKone et al. (2016) from discussing parameter estimates and power analysis.

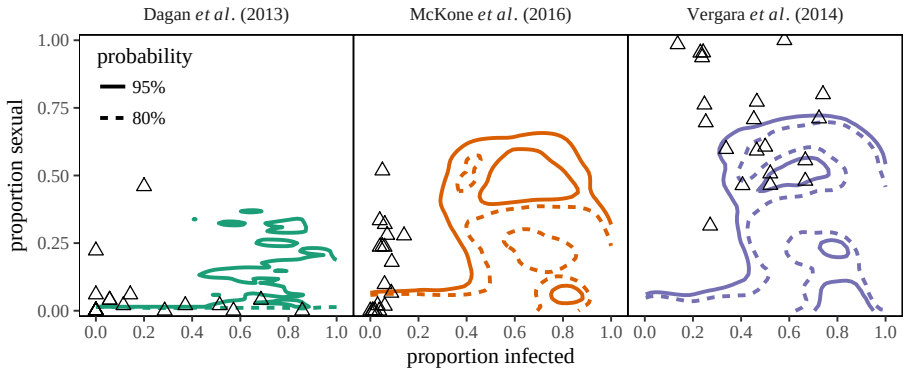

Figure 2: **Comparison in trends of the observed data v.s. simulated data from the posterior samples.** Solid circles represent a population within a simulation. Triangles represent observed data.

The most striking result from Fig. 2 is that there is a region (around 30%-40% infection prevalence) in which proportion of infected hosts decreases while proportion of sexual hosts increases (most clearly visible in the fits to McKone et al. (2016) and Vergara et al. (2014b)). As transmission rate (β) increases, selection for sexual hosts increases but increasing number of resistant offsprings prevents further infection from occuring and can decrease overall infection preva-

lence. Such trend is consistent with the results of Lively (2001) who noted that there is an overlap in infection prevalence that selects for sexual and asexual reproduction. We also find that proportion of sexual hosts decreases when infection prevalence is very high. Decrease in fitness of sexual hosts associated with increase in prevalence was predicted by Ashby and King (2015); it can also be found in an earlier work by Lively (2010b) although it was not discussed in the paper.

Parameter estimates for Vergara et al. (2014b) are presented in Fig. 3. The most surprising result is that the posterior distributions of the scale parameter for cost of sex, c_b, is much wider than the prior distribution and have noticeably higher mean: (mean (95% credible interval)) 1.30 (0.80-1.97). 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 corresponds to the following mean and 95% CIs for cost of sex: 1.54 (1.02-2.50). Note that some posterior samples range $c_b > 2$, corresponding to faster growth rate of sexual hosts than asexual hosts. Simulated data from these posterior samples mostly consist of populations with almost 100% sexual hosts. While these estimates are not entirely impossible given that Vergara et al. (2014b) observed almost 100% sexual snails in one of their sampling sites throughout 5 years, their observation is likely to be a sampling artifact since an earlier work sampled at the same site reported much lower frequency of sexual snails (Vergara et al., 2013).

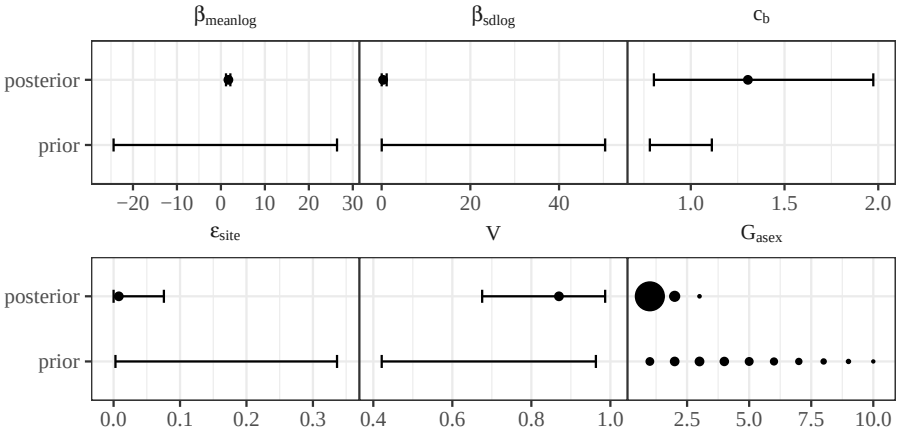

Figure 3: **Parameter estimates from Sequential Monte Carlo Approximate Bayesian Computation.** Lines repreesnt 95% quantile and points represent mean posterior estimates. 100 posterior samples were obtained from SMC ABC. Prior distributions are specified in [TODO].

The model estimates high virulence overall: (mean (95% CI)) 87.0% (67.5%-98.7%). In contrast to Lively (2010b), who modeled 1 asexual genotype com-

peting with 9 sexual genotypes, our model estimate shows that higher asexual to sexual genotypic ratio can be supported (G_{asex} panel in Fig. 2). 274 out of 300 posterior samples estimate $G_{\text{asex}} = 1$; 25 samples estimate $G_{\text{asex}} = 2$; and 1 sample estimates $G_{\text{asex}} = 3$. However, we find that it is still necessary for sexual hosts to have higher genetic diversity than asexual hosts.

Finally, a power analysis show that the power for detecting a positive correlation between infection prevalence and frequency of sexual hosts in Vergara et al. (2014b) population is almost 0 (Fig. 4). We find that there is much higher (but still low) power to detect a negative correlation. Increasing number of samples per site has small effect on power once the sample size greater than 50. Increasing number of sites leads to greater increase in power but the power appears to saturate as number of sites increases.

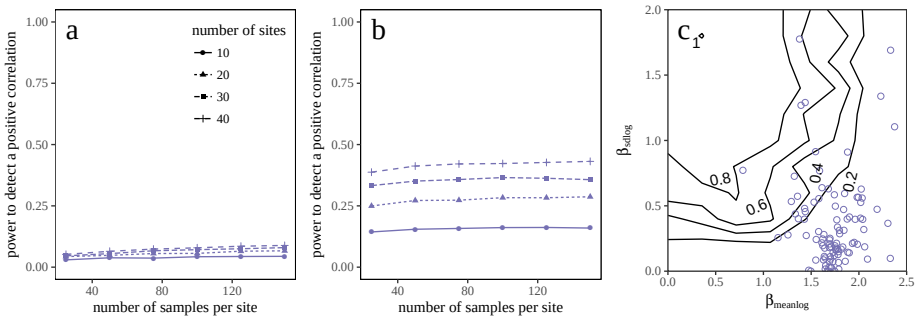

Figure 4: **Power to detect a statistically clear correlation between infection prevalence and frequency of sexual hosts.** (a) Power to detect a positive correlation. (b) Power to detect a negative correlation. (c) (Contour) power to detect a positive correlation as a function of transmission rate parameters and (points) marginal posterior distributions. 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. Contour plot is created by taking the mean posterior estimates and assuming 20 sites and 100 samples per site.

4 Discussion

Our results challenge ways in which the Red Queen Hypothesis for sex has been studied. Previous modeling studies have either relied on assumed parameters [CITE] or explored parameter spaces [CITE] to understand sexual reproduction maintained by host parasite coevolution. Here, we show that (1) model parameters are estimable and (2) a simple Red Queen model with population structure can reproduce key summary statistics observed across three different snail systems (Fig. 1). While our model is able to reprouce observed summary statistics, there still remain some discrepancies between model prediction and

observed relationship between infection prevalence and frequency of sexual hosts across populations. These discrepancies suggest a simple host-parasite model is not sufficient to explain sexual reproduction observed in nature.

A model that does not fit well can sometimes tell us more about the system of interest than a model that fits well. For example, there is a clear mismatch between the model and the data presented by Dagan et al. (2013a) (Fig. 2). The snail populations studied by Dagan et al. (2013a) live in intrinsically different environments from two other snail populations that we consider. For example, some habitats are subject to seasonal flash floods, which can affect reproductive strategies of snails (Ben-Ami and Heller, 2007). Althuogh we cannot estimate the strength of the effect of environment on sexual reproduction of snails relative to that of the effect of parasites, it is likely to be a cause of the bad fit. We conclude that the Red Queen Hypothesis alone cannot explain maintenance of sexual reproduction in this population.

The only system that our model could resemble fairly well is one studied by Vergara et al. (2014b). In order for our model to produce similar summary statistics as those observed in nature, it requires the scale parameter (c_b) for the cost of sex to have mean greater than 1 (Fig. 3). A model that assumes two fold cost of sex ($c_b = 1$) would not have fitted well. However, from an inferential point of view, $c_b > 1$ does not necessarily imply that cost of sex is less than two fold. We interpret an estimate of $c_b > 1$ as amount of compensation required in order for the model to reproduce observed data. Additional compensation can include less than two fold cost of sex or any other mechanisms that can promote sex.

A suitable candidate for model compensation is increase in host genetic diversity. Although exact genetic architecture that determines trematode infecion in snails (e.g., loci involved in parasite resistance) is not known CITE, genetic diversity of snails that have been documented is far greater than what we have assumed (King et al., 2011; Dagan et al., 2013a). In addition, increasing genetic diversity of the model may resolve overestimation of infection prevalence (Fig. 1). Previous studies have shown that moderately high genetic diversity can allow sexual hosts to escape infection more easily and therefore reduce prevalence of infection given similar amount of sexual reproduction (Lively, 2010a; King and Lively, 2012; Ashby and King, 2015). Overall, our results indicate that more modeling effort is required to understand prevalence of sexual reproduction in nature.

Our power analysis (Fig. 4) constrasts with the positive correlation predicted by Lively (1992, 2001) and findings of many empirical studies that have confirmed the prediction (Lively, 1987; Lively and Jokela, 2002; Kumpulainen et al., 2004; Vergara et al., 2013; McKone et al., 2016). The power analysis predicts almost no power for detecting a positive corrleation in the population studied by Vergara et al. (2014b) and relatively higher but still low power for detecting a negative correlation. This result may appear to contradict an earlier work by Vergara et al. (2013) that reported a positive correlation between between infection prevalence and male frequency in the same lake but there is a simple explanation for the difference. The key premise behind the postive

correlation predicted by Lively (2001) is that there must be large variation in infection prevalence. In particular, range of prevalnece must be wide enough so that the sample includes sites with almost no infected hosts (hence no sexual hosts) and those with reasonably high proportion of infected hosts to maintain sexual reproduction through parasitism (Lively, 2001). Since all four habitats studied by Vergara et al. (2014b) consists of populations with high prevalence and high frequency of sexually reproducing hosts, positive correlation vanishes. Instead, studying a system with larger variation and lower mean prevalence will yield much high power (Fig. 4(c)).

On the other hand, negative correlation between prevalence of infection and frequency of sexual hosts can be explained by cycling of host and parasite populations. A main component of the Red Queen Hypothesis is that negative frequency dependence drives oscillation in both host and parasite population (Hamilton, 1980). When temporal variation is taken into account, association between infection prevalence and frequency of sexual hosts can change depending on what phase each of the sample population is going through in its cycle (Fig. 5). The negative correlation in the population does not contradict the positive correlation predicted by Lively (2001) because their prediction did not take temporal variation into account.

References

Agrawal, A. and C. M. Lively (2002). Infection genetics: gene-for-gene versus matching-alleles models and all points in between. *Evolutionary Ecology Research 4*(1), 91–107.

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

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.

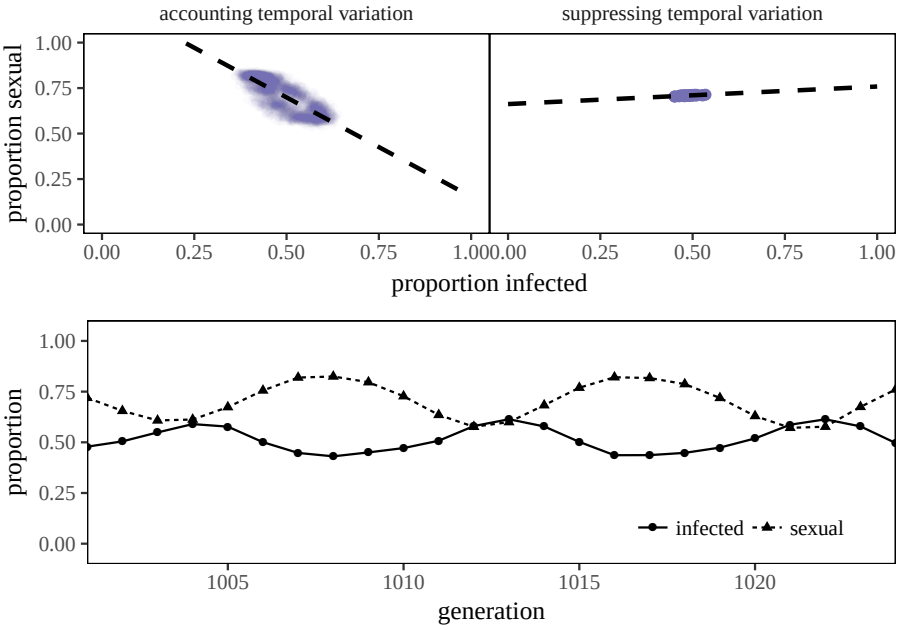

Figure 5: **Simulated data from the posterior distribution fitted to Vergara et al. (2014b).** A particular simulation is chosen from the posterior distributions to demonstrate that it is possible to predict opposite relationship when temporal variation is taken into account. (Top left) each point represents proportion of infected hosts and proportion of sexually reproducing hosts of each population at each generation. Last hundred generations are plotted. (Top right) each point represents mean proporiton of infected hosts and mean proportion of sexually reproducing hosts of each population averaged over the last hundred generations. Dashed lines represent least squares fit to all point. (Bottom) A typical host parasite cycle observed in a population from this simulated data.

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

Dybdahl, M. F. and C. M. Lively (1995). Host–parasite interactions: infection of common clones in natural populations of a freshwater snail (*Potamopyrgus antipodarum*). *Proc. R. Soc. Lond. B 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.

Galvani, A. P., R. M. Coleman, and N. M. Ferguson (2003). The maintenance

of sex in parasites. *Proceedings of the Royal Society of London B: Biological Sciences 270*(1510), 19–28.

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

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.

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

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

MacPherson, A. and S. P. Otto (2017). Joint coevolutionary-epidemiological models dampen red queen cycles and alter conditions for epidemics. *Theoretical population biology*.

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.

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.

Otto, S. P. (2009). The evolutionary enigma of sex. *The American naturalist 174*(S1), S1–S14.

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

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

Tobler, M. and I. Schlupp (2008). Expanding the horizon: the red queen and potential alternatives. *Canadian Journal of Zoology 86*(8), 765–773.

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.

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