Quantifying the effects of parasites on the maintenance of sex

Sang Woo Park, Benjamin M Bolker

November 30, 2017

1 Introduction

Sexual reproduction is followed by numerous costs (Lehtonen et al., 2012), and these costs are often attributed to cost of producing males (Smith, 1978). As males cannot produce offspring, a sexually reproducing population should be outgrown by an asexually reproducing population growing twice as fast (given that the sexual population produces 50% males and 50% females) – hence the infamous *two-fold cost of sex* (Smith, 1978). However, the two-fold cost of sex relies on the assumption that all else is equal, and this is not necessarily true in nature. In fact, sexual reproduction is much more common in nature (Vrijenhoek, 1998). So what drives sex to persist?

One explanation for sex is the Red Queen Hypothesis (Bell, 1982). The Red Queen Hypothesis predicts sexually reproducing organisms to overcome the two-fold cost of sex under strong parasite selection by producing genetically diverse offspring that are resistant to infection (Haldane, 1949; Jaenike, 1978; Hamilton, 1980). Hosts with rare genotypes are expected to result in higher fitness and become common eventually. Furthermore, parasites coevolves with the host population to infect the common "rare" genotype [CITE], and negative frequeny dependnece drives an oscillation in genotypic frequencies in host and parasite populations (Clarke, 1976; Hamilton, 1980).

Since the Red Queen Hypothesis has been postulated, much of theoretical work has focused on figuring out conditions under which parasites can maintain sexual reproduction.

On the other hand, empiricists have tried to quantify the Red Queen.

While the Red Queen Hypothesis is well supported both empirically and theoretically, 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 (Ashby and King, 2015). For example,

Recent studies have shown that ecological feedbacks can allow sex to be maintained more easily [CITE]. We find that a simple model can capture the dynamics Here, we try to bridge this gap further.

In this study, we extend the model suggested by Lively (2010b) to include stochasticity and population structure and fit the to data from Dagan et al.

(2013); McKone et al. (2016); Vergara et al. (2014) using approximate bayesian computation (ABC). [Something about what we found out from the model fitting] Using posterior distributions obtained from the fit, we perform power analysis to test the idea that infection prevalence is positively correlated with frequency of sexual reproduction Lively (2001). We find that cycling of parasites and hosts can lead to negative correlation between prevalence of infection and sex. Instead, positive correlated predicted by Lively (2001) is driven by populations that do not go through Red Queen dynamics (i.e., having low level of sexual and infection).

2 Methods

2.1 Data

We consider two snail-trematode population in New Zealand (McKone et al., 2016; Vergara et al., 2014) and a similar snail-trematode population in Isral (Dagan et al., 2013). First, the snail-trematode system in New Zealand has been extensively studied under the Red Queen Hypothesis [CITE] so we expect a Red Queen model to fit well. Second, [Something about Israel snails being different]. Finally, by considering both spatial (Dagan et al., 2013; McKone et al., 2016) and spatiotemporal data (Vergara et al., 2014), we can compare how structure of the data may affect model fits. Data for Vergara et al. (2014) and Dagan et al. (2013) were obtained from their Dryad repository [CITE] and data for 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 (reduction in 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 population dynamics models and is general enough to be applied to broad systems in nature. We do not consider mechanistic details of the snail-trematode system such as life history of trematodes (Vergara et al., 2014) as it would require us to estimate or make assumptions on parameters on mechanisms, both of which are unrealistic.

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 $A_{ij}(t)$ unless necessary. 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, $S_{ij,U}$ and $S_{ij,I}$ are the number of uninfected and infected sexual hosts in a population, and W_U and W_I are their corresponding fitnesses. We allow for cost of sex to vary by using 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 that there is mixing between population, determined by rate ϵ_{mix}. Then, the expected number of offspring in the next generation (accounting for contributions from all populations) is given by

$$\begin{aligned}
\mathrm{E}\left(S^k_{ij}(t+1)\right) &= f_{\text{sex}}\left((1-\epsilon_{\text{mix}})\left(S^k_{ij}\right)' + \frac{\epsilon_{\text{mix}}}{n_{\text{population}}-1}\sum_{l\neq k}\left(S^l_{ij}\right)'\right), \\
\mathrm{E}\left(A^k_{ij}(t+1)\right) &= (1-\epsilon_{\text{mix}})\left(A^k_{ij}\right)' + \frac{\epsilon_{\text{mix}}}{n_{\text{population}}-1}\sum_{l\neq k}\left(A^l_{ij}\right)',
\end{aligned} \tag{3}$$

where $f_{\text{sex}}(x)$ is the function that models sexual reproduction, including recombination at a rate r_{host} and outcrossing, and $n_{\text{population}}$ is the number of populations present in the system. We then take a poisson random variable to simulate process error and allow for stochastic migration to avoid fixation:

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

where $p_{ij,\text{sex}}$ and $p_{ij,\text{asex}}$ 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 genotypes

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 $S_{ip,i,I}$ and $A_{ip,i,I}$ are the expected number 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 $p_{i,\text{parasite}}$ 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. 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

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

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

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

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

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

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 with haplotype 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 (see (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 migration of asexual hosts with genotypes genotypes but fix number of asexual genotypes (denoted by G_{asex}) that can be present in the population, while the number of sexual genotypes (G_{sex}) that can be present in the population is equal to 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

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

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. 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 can be particularly useful when the exact likelihood function is not available [CITE?]. 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.

For each sample from the posterior, the model is simulated and a subsample from the simulated data is drawn so that the number of simulated population is equal to the number of sites collected in a study. Then, the summary statistics are calculated by considering last 100 generations of simulated data and the parameter sample as well as the simulation generated by the sample is accepted if the distance between simulated and observed data (measureed by the absolute sum of differences in summary statistics) is less than the tolerance value. By simulating greather number of populations than the observed, we account for uncertainty in unobserved population as well as interaction between populations.

We use uninformative priors for all parameters except c_b, a scale parameter for the cost of sex. 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) (see ??? for the list of

parameters). As Dagan et al. (2013) 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 Sequential Monte Carlo Turner and Van Zandt (2012) (see Appendix for details).

3 Results

3.1 Model fit

Before we present the parameter estimates, we first discuss the resulting fits of the model and their goodness of fits. When we compare the summary statistics of the observed data and those estimated by the model, it appears that our simple Red Queen model can reproduce observed data reasonably well (Fig. 1). The only discrepancy is that the model overestimates prevalence of infection, given similar level of sexual reproduction. In particular, the data presented by McKone et al. (2016) has mean infection prevalence of approximately 5% but the model predicts prevalence of approximately [TODO].

High estimate of infection prevalence can be explained by limited genetic diversity of the model. Previous studies have shown that moderately high genetic diversity can allow sexual hosts to escape infection more easily and therefore reduces prevalence of infection while maintaining sexual reproduction (Lively, 2010a; King and Lively, 2012; Ashby and King, 2015). However, our model assumes that infection is deteremined two biallelic loci. Although not much is known about exact loci that determine trematode infecion in snails, it is not likely that a two loci model is sufficient to capture the genetic diversity of snails [CITE]. Therefore, we expect a model with limited genetic diversity to have higher infection prevalence than observed prevalence.

As comparing summary statistics only present marginal fits of the model, we also compare mean infection prevalence and mean proportion of sexually reproducing hosts (averaged over last 100 generations) of each simulated population along with the observed proportions (Fig. 2). The most striking result from Fig. 2 is the negative correlation between prevalence of infection and frequency of sexually reproducing hosts in the intermediate range of infection prevalence (most clearly seen from the fits to Vergara et al. (2014)). As infection rate increases, frequency of sexual reproducing hosts increases but increasing number sexual hosts in the population decreases prevalence of infection as more hosts are escaping infection. Our result contrasts with the work by Lively (2001) who predicted positive correlation in prevalence of infection and frequency of sexually reproducing hosts.

We also find that frequency of sexual hosts decrease when prevalence of infection is very high (near 100%). When genetic diversity is limited and infection rate is high, sexually reproducing hosts fail to escape infection and sexual reproduction cannot be maintained Ashby and King (2015). This trend was also observed in Lively (2010b), which we used as a basis of our model, but was not

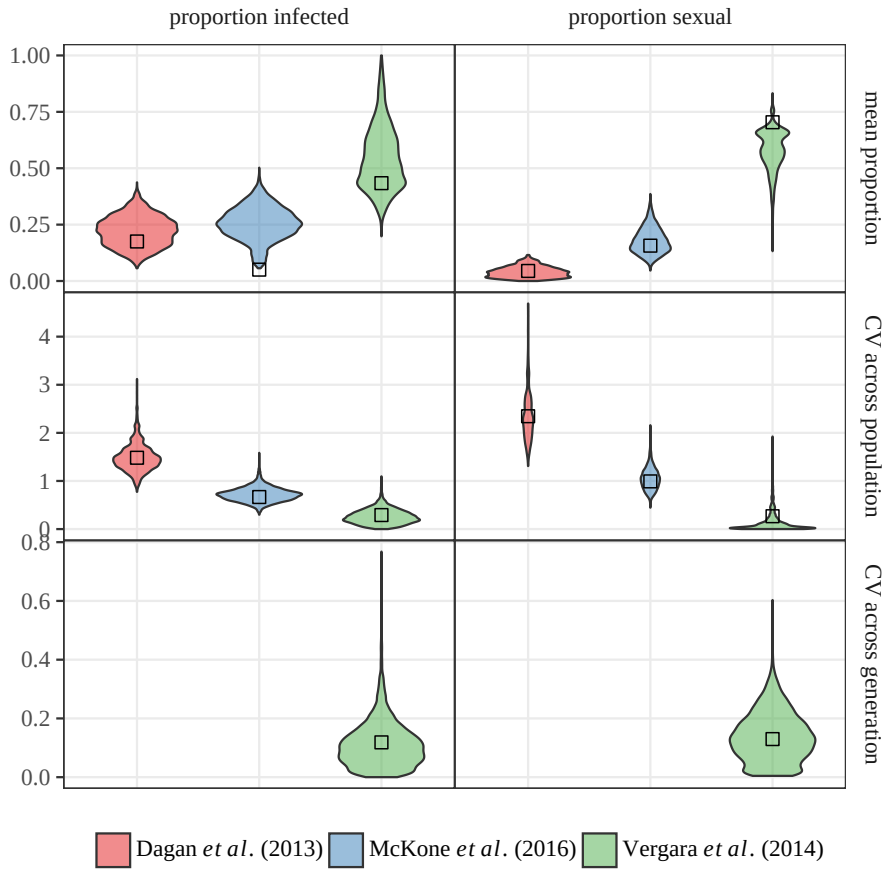

Figure 1: **Summary statistics of the observed data v.s. simulated data from the posterior samples.** Violin plots show distribution of summary statistics from each simulated set. For each simulated data set accepted along with the posterior parameter. a set of simulated populations are sampled at random 100 times and summary statistics are calculated for each bootstrap sample by considering last 100 generations. Squares represnt observed summary statistics.

discussed.

Fig. 2 may seem to suggest the model fits are much worse overall than what we expected by looking at Fig. 1. It is important to note that we are plotting the mean proportions (suppressing the effects of the cycle) whereas observed data is more likely to be traces of the Red Queen cycles.

Yet, we conclude that our model does not fit to Dagan et al. (2013). Simulated data from the model is able to reproduce summary statistics of the data

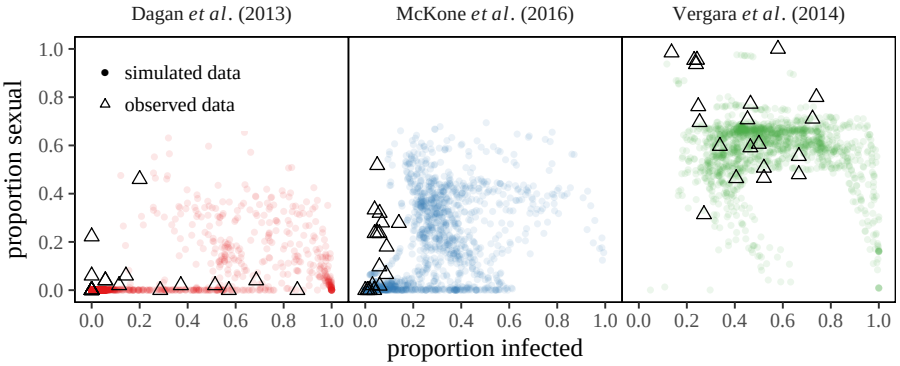

Figure 2: **Comparison in trends of the observed data v.s. simulated data from the posterior samples.** Solid circles represent a population within a simulation. Triangles represent observed data.

observed by Dagan et al. (2013) (Fig. 1) but shows a completely different trend in the maintanence of sexual reproduction across infection prevalence (Fig. 2). Our model predicts sexual reproduction to be maintained at intermediate infection prevalence, but the data presented by Dagan et al. (2013) show that sexual reproduction is only maintained at a low level of infection prevalence and is barely maintained otherwise.

While all three studies (Dagan et al., 2013; McKone et al., 2016; Vergara et al., 2014) look at snail-trematode systems, the system studied by Dagan et al. (2013) live in intrinsically different environments. [TODO: add details] Our model not take any environmental factors into consideration and only depends on parasite mediated selection for sex. It is not expected that our model should fit well. Due to such badness of fit, we exclude fits to Dagan et al. (2013) from further analyses.

3.2 Parameter estimation

Our model estimates wide ranges of parameters across data sets (Fig. 3). The most surprising result is that the posterior distribution of the scale parameter, c_b, estimated by the model is much wider than the prior distribution. In most cases, posterior distribution is expected to be narrower than prior distribution or at most similar to prior distribution when data provides minimal or no information about a parameter of interest [CITE?].

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 95% credible intervals for cost of sex: [???]. However, the meaning of the parameter changes from an inferential point of view.

In the absence of the parameter c_b (i.e, when $c_b = 1$), our model assumes two

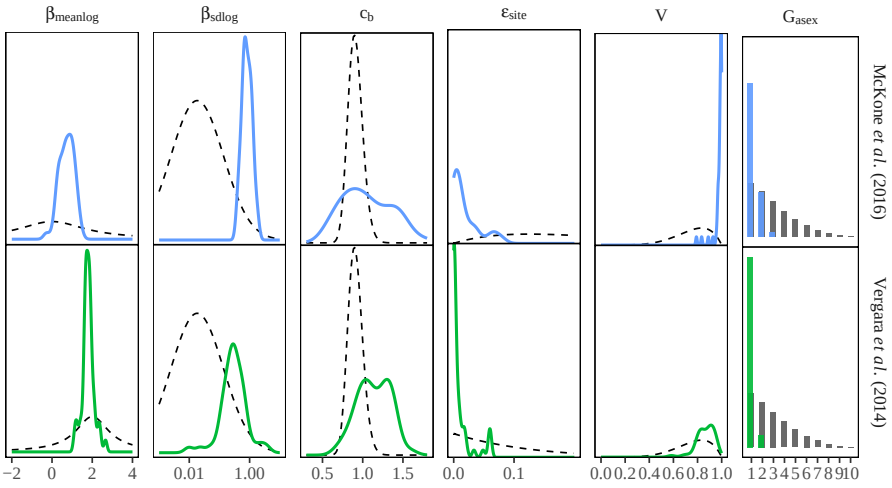

Figure 3: **Need caption.** Need caption

fold cost of sex in growth rate. Then, an estimate of $c_b > 1$ can be interpreted as the amount of compensation required in order to overcome the two fold cost of sex. Additional compensation may include other mechanisms that helps maintenance of sex, such as higher genetic diversity.

Similarly, an estimate of $c_b < 1$ implies that the model can sufficiently produce observed data without any compensation.

We also find that the model estimates high virulence overall and almost 100% virulence for the system studied by McKone et al. (2016).

3.3 Power analysis

Power analysis reveals that detecting a posi

4 Discussion

Here, we find that a simple Red Queen

References

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

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

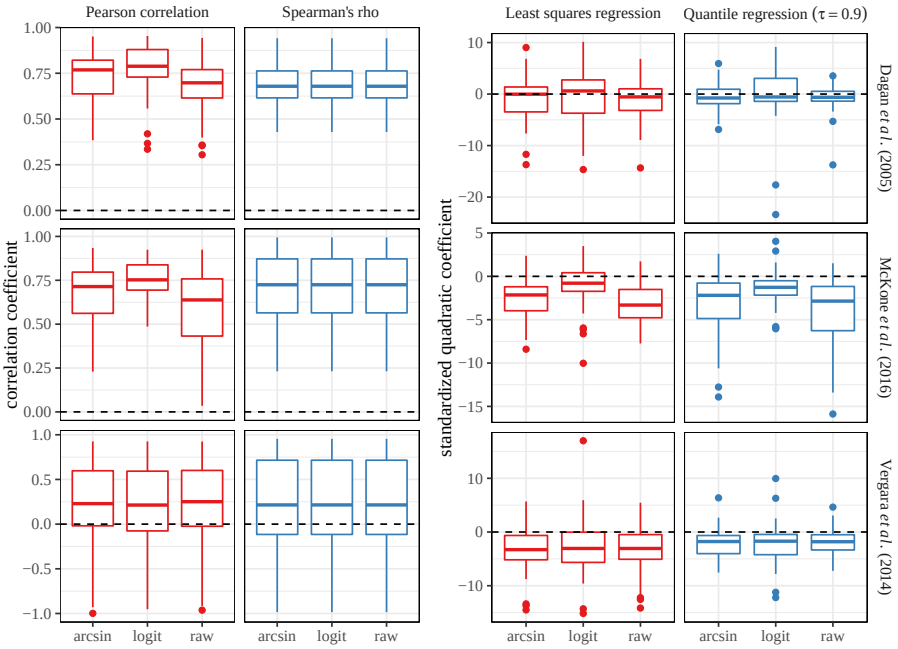

Figure 4: **Need caption.** Need caption

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 (2013). Clonal diversity driven by parasitism in a freshwater snail. *Journal of evolutionary biology 26*(11), 2509–2519.

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.

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.

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

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

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

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.

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. 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.

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

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