Host-parasite evolution in male-haploid hosts: an individual based network model

J. Kidner · Robin F. A. Moritz

Abstract Host-parasite co-evolution is a key component of the Red Queen Hypothesis (RQH). The RQH currently being one of the main hypotheses describing the evolution of sex and recombination. However, most analyses in this area have either ignored parasite transmission or included it either with mean field or simple frequency based models. Moreover models have rarely addressed the issue of male haploid species. We here use agent based models to qualify the interactions between host- and parasite-based transmission parameters and virulence comparing diploid with male-haploid species. We found diploid hosts to have a higher fitness under the inverse matching allele mode compared to male haplodiploid hosts which in turn have a higher fitness under the matching allele model . Selection for recombination was rare but whenever selection for recombination was evident (<6.6 %), the resulting recombination rates were both consistently higher and more frequent in male haploids.

Keywords Red Queen Hypothesis · Hymenoptera · Network models · Coevolution

Introduction

Evolutionary theories explaining the development and maintenance of sex (those involving meiosis) frequently employ the Red Queen Hypothesis (RQH). The RQH is derived from

J. Kidner (✉) · R. F. A. Moritz
Fachbereich Biology, Institute of Biology, Martin-Luther-University, Halle/S., Sachsen-Anhalt, Germany
e-mail: jonathan.kidner@zoologie.uni-halle.de

R. F. A. Moritz
Department of Zoology and Entomology, Social Insect Res Grp, University of Pretoria, Pretoria 0002, South Africa

Lewis Carroll's character, the 'Red Queen', in Through the Looking Glass with the comment 'it takes all the running you can do, to keep in the same place' (Carroll 1960). The analogous form in population genetics is that it takes continuous selection (from parasites/pathogens) and fluctuations in the host fitness landscape (as parasites increase/decrease in frequency) to maintain selection for sex, and/or recombination (Sutton et al. 2011; Takahashi et al. 2011). The measure used to predict recombination's selection landscape is a combination of linkage disequilibria (LD) and Epistasis (symbolised together by $LD \times E$). If the host fitness landscape changes infrequently then the predominant selection landscape will be against recombination (as recombination will breakdown favourable allele associations). Barton (1995) predicted a need for a period of 2–5 host generations before $LD \times E$ changes direction, but Salathé et al. (2007) showed that this requirement can be substantially relaxed.

While the RQH is primarily developed to explain the occurrence of sex and/or recombination, as is the emphasis here. The envisioned process (negative frequency dependent selection) has a strong influence on maintaining genetic diversity. However the main impact on genetic diversity will come from differences in effective population size. For haplodiploid organisms, or X-/Z-chromosomes the effective population size is $3/4$ that of diploid chromosomes. This having an impact on both standing diversity from genetic drift and the efficacy of selection. Either of these processes can be ameliorated through increasing recombination.

Analyses of the RQH have been limited both by few levels of ploidy (typically haploid and diploid) and adoption of three or less interaction models. In general ploidy has been ignored as model factors (Agrawal and Lively 2002; Nuismer and Otto 2004; Otto and Nuismer 2004; Peters and Lively 1999, 2007; Schmid-Hempel and Jokela 2002). Those studies that compared the ploidy of hosts on the RQH are few and far between (Nuismer and Otto 2004; M'Gonigle and Otto 2011; Kidner and Moritz 2013) with only Nuismer and Otto (2004) considering ploidy amoung parasites. Nevertheless, the predictions made for the host species amoung these studies (with different model designs) remain remarkably consistent. In contrast, comprehensive comparisons of the interaction models are currently limited to just a single paper (Engelstädter and Bonhoeffer 2009) which also suggests flaws in the assumption that oscillations in $LD \times E$ are necessary for the evolution of recombination.

A further constraint in the current theoretical work on the RQH is on parasite transmission. Most studies make the 'single challenge' assumption: Every host individual is considered to be 'challenged' by a single parasite (Agrawal and Lively 2002; Agrawal 2009; Hodgeson and Otto 2012; M'Gonigle and Otto 2011; Peters and Lively 2007). The single challenge assumption disregards both experimental and theoretical studies on parasites/pathogens. Lively (2010) developed a mean field transmission model within their analysis of the RQH. showing the importance of parasite growth (R_0) for the RQH (Kidner and Moritz 2013; Lively 2010). However, several restrictions of the mean field model have been recorded (Boots and Mealor 2007; Webb et al. 2013; Keeling 1999) from epidemiology. Mean field models cannot replicate results from empirical studies or network models when parasite/pathogen presence and transmission are low (Keeling 1999). They also cannot lead to insights gained from models/simulations concerning parasite/pathogen evolution (Boots and Mealor 2007; Webb et al. 2013).

The aim in the current study is to assess the potential importance of network structure on the RQH. While theoretical studies over ecological timescales have demonstrated a role for networks, this has yet to be done over evolutionary timescales. Furthermore, employing the use of an agent based model (ABM) allows the adoption of a genetic algorithm (GA) to

assess the recombination rate. This GA uses the implicit selection from parasite infections on those individuals with unfavourable genetic combinations. These assessments will be done for both the matching allele model (MAM, infections occur when parasite genotypes match the host) and the Inverse MAM (IMAM, infections occur when parasite genotypes mismatch the host). In addition to this we will compare diplo-diploid with haplo-diploid (male–female) systems because the latter have been shown to have the highest recombination rates-in the animal kingdom (Meznar et al. 2010; Stolle et al. 2011; Wilfert et al. 2007). The most divergent haplodiploid lineage is that of the Hymenoptera, an order that contains some of the most complex social systems with very tight network structures (Normark 2003). To approximate the social networks that occur within the Hymenoptera we use two measures: the degree of locality (cliqueness), and the average number of connections per node (mean degree). While some measures of these are available from small network studies of the Bumblebee (Otterstatter and Thomson 2007), we also randomize these values to provide a broader overview.

In the following simulation we develop a network with cliqueness and connectedness parameterized. In addition each node contains genetic information and a personal attribute indicating the recombination rate. From the parasites the rate of infection along a connection, virulence (cost of infection to the host) and the parasite-host generation ratios are parameterized. The model is then analysed for these parameters both visually and with GLMMs from the lme4 package (R Core Team 2013).

Materials and methods

The simulation was based on a two locus two allele model. Haploid individuals were treated as their equivalent homozygous diploid genotype. The parasites were treated as clonal lineages, the parasite driven processes were considered to be independent between strains. The two loci interaction matrix was setup according to the MAM and IMAM (see Supplementary material Table s1). With an ABM, an GA approach was used for estimating evolutionary stable states for the recombination rate. All randomization procedures and generation used the mersenne twister mt19337b algorithm from the C++ booster packages.

Network construction

We used an agent based network model inspired by the work of Keeling (1999) and others (Ames et al. 2011) in epidemiology (Fig. 1). The algorithms used for construction were similar to those used for scale-free networks, but with diminishing returns on the connection probability density. Network topology was described by the mean degree (which was allowed to vary from 2 to 6 edges/connections) and cliqueness (which ranged from 0.05 to 0.90). Cliqueness describes the degree to which an edge is shared between the first and last nodes of a chain of three: Of the nodes a–b–c a direct connection between nodes a–c would form a clique.

Each simulation run uses a standardised set of individuals in the first host generation (uniform genotype distribution and recombination). In subsequent generations hosts are created from the gametes of two randomly chosen parents (amoung the previous generation). After construction of a new host population (algorithms 9 11, supplementary material 1) the host network was constructed (algorithms 1 3, supplementary material 1). Inclusion of the first and second nodes requires the addition of a single edge between these two

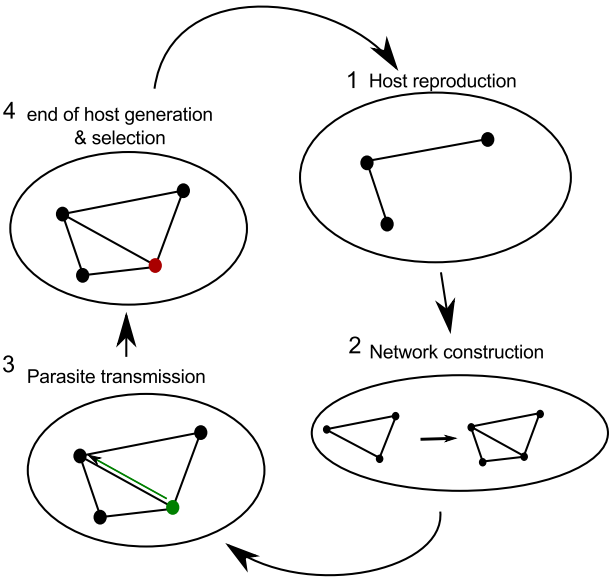


Fig. 1 The basic model layout, split into four parts. (*1*) The first required stage is to set up the host population. This is done by simulating host reproduction with a uniform systematic selection of host genotypes. (*2*) Afterwards, the host network is constructed using a scale-free algorithm for adding nodes. After this the network is then refined to meet the specific network parameters (cliqueness and connectedness). (*3*) Upon construction of the network at least one individual per strain will be selected for infection. Upon which transmission is then allowed between multiple parasite generations. (*4*) On the last parasite generation selection coefficients against the various host entites are calculated. Those surviving entities are then used to generate the next host generation, upon which step (*1*) is entered

nodes. Addition of a third node may require a choice to be made if only a single neighbour is to be chosen. After the addition of $n_{min} + 1$ nodes to the network neighbours are randomly assigned from at least n_{min} earlier nodes. Assignment of the neighbours is based on a probability density, the more neighbours a node has the more likely they are to create new neighbours. Each time a node gets a new neighbour its probability density (p) is modified using $(p = \sum_1^n \frac{1}{\sqrt{n}}$, n is the new number of neighbours).

Upon addition of the final node in the network, the cliqueness of the network was adjusted by adding new edges between neighbouring hosts (Fig. 1 step 2). The new edges were added to triplets that did not form cliques, on addition of an edge to form a new clique, a separate edge was removed. The edge for removal was chosen from the same group of nodes where the new clique was formed. With virulence being measured as parasite induced prevention of host reproduction (sterilization), host networks remained static per host generation. Virulence was varied from low (0.05) to high (0.95) degrees.

Host populations

Infected hosts could survive and reproduce according to the specific parasite virulence. Hence surviving infection *(s)* was calculated as 1 minus the mortality rate when the parasite strains were present $(s = \prod_{i=1}^{4} P_i * (1 - Parasite\_virulence))$. P indicates the

vector of presence/absence of the individual parasite strains (0 for absence and 1 for presence). To enter into the breeding pool a random draw based on the survival probability (s) was used. Due to the possibility of skewed sex ratios (from differences in selection between the sexes), individuals were drawn at random without replacement from the breeding pool. The offspring genotypes were then generated from this subset of breeding individuals with the parental recombination rates being taken into account.

On initialization of the simulation each individuals recombination rate was set to 0.5, and the decay of these values was subsequently measured. With the different rates of selection based on host genotype, also determined by the recombination rate, a GA can be used to analyse the RQH.

Parasite transmission

All of the different parasite genotypes were assumed to act independently of each other simplifying both the infection algorithms and calculations. In the first host generation ($t = 0$) infections were limited to one host individual per parasite genotype. In subsequent generations the specific 'novel infections' were chosen at random from the susceptible population.

For each parasite strain three host states exist: susceptible, infected and resistant. Resistance and susceptibility were determined according to the genotypes and interaction model (MAM, or IMAM, see Table s1). Because resistance was determined by genotype, the resistant state was a static parameter for each parasite strain. In contrast the susceptible state could transition to infected, but could not recover to either a resistant or susceptible state. A consequence of this is that parasite transmission then becomes reduced to three parameters: the number of infected (I) susceptible (S) pairs, the transmission between such pairs, and the host-parasite generation ratio. The host-parasite generation ratio was determined as ranging from slow (2 parasite generations per host generation), to fast (12 parasite generations per host generation).

Transmission from parent to offpsring generation used a horizontal approach. Conceptually the approach used would be like a parasite 'spore bank'. Parasites from the parent generation provided a resource within the environment that could infect the offspring generation. With this approach transmission was calculated by the number of infected multiplied by the frequency of susceptible individuals in the offspring generation. In every host generation the parasite population was not allowed to fall below unity. If no susceptible individuals were available then the simulation was terminated and recorded as not supporting the RQH. Results from these simulations were only retained if the simulations ran for at least 1,000 host generations before extinction of a susceptible population.

Analyses

The population size was set to 4,000 and the number of host generations to 5,000. These values were used to reduce the impact of stochasticity on the host, while also reducing computational time. 5,000 host generations was chosen after running of the simulation (after 3,000–4,500 generations, the recombination rate was not recorded to change). For analysing host-parasite dynamics we collected host and parasite counts sorted by genotype. For moving through parameter space we collected the: degree of clustering (cliqueness, 0.05–0.95), connectedness (2–6), virulence (0.05–0.95), I–S infection rate (0.05–0.8), recombination rate (initially set to 0.5), the Vmax for recombination and the model stability (whether host alleles went extinct, or not). Two factors were then added: a boolean

for a diploid, or haplodiploid host; and a boolean for the use of either a MAM, or IMAM. Collection of the Vmax of recombination rate was through measuring the change in recombination averaged over 50 generations.

All analyses were performed in either R (packages vcd, nlme, lme4; R Core Team 2013), or GNUplot (Williams and Kelly 2003) to test predictions about the frequency dynamics of haplotypes or alleles. GLMM models were constructed through reduction of non-significant parameters from an initial model of all parameters and first order interactions (model simplification according to Crawley 2005).

Results

Interactions between interaction model and ploidy

From the individual simulation runs haploid individuals tended to have lower susceptible host frequencies than diploids under the MAM (Fig. 2). This has two major consequences: Firstly, the overall frequency of susceptible hosts is higher in the diploid population (from an average of 0.148–0.160). Secondly, the lower frequency of susceptible host genotypes reduces the prevalence of parasite strains. Inverting the interaction model (the lower 2 graphs of Fig. 2) reverses this relationship between host ploidy and susceptible frequencies (from −0.171 to −0.148; difference between diploid and haplodiploid host populations).

Parasite growth

The second consequence of higher susceptible frequencies is the improved chance of establishment within the host population (grey lines in Fig. 2). To test the importance of this up to 100 complete waves (beginning and ending at peaks of susceptible frequencies) were sampled randomly per haplotype over 4,000 generations. The median peak height per parasite strain (number of infected) in diploids (under the MAM) was recorded as 450 out of a population of 5,000 (see supplementary material Table s2). In haplodiploid hosts the median was 409.5 (under the MAM). Under the IMAM the inverse was observed, though at much lower levels (169.5 infected in haplodiploid hosts, compared to 124.5 in diploid). Parasite populations follow the pattern of abundance of susceptible host frequencies. The median period length of the parasite populations tends to remain much higher than the range hypothesised to be necessary (Barton 1995; Peters and Lively 2007).

Host fitness

The relationship between fitness of the host and interaction model follows that of parasite prevalence (Fig. 3) . Host fitness is higher for haplodiploids under the MAM, the converse being true under the IMAM. The functions translating parasite virulence to fitness of the host are quadratic under all conditions. Though the exact functions are different between conditions, with gentler troughs occurring for the fitness of haplodiploid hosts. Under all scenarios the lowest average fitness is recorded (per host) under those conditions in which we would estimate the highest proportion of I–S edges. It is also under these cases that we would expect to see the largest differences in fitness between males and females of the haplodiploid population (according to the masking hypothesis Otto and Marks 1996).

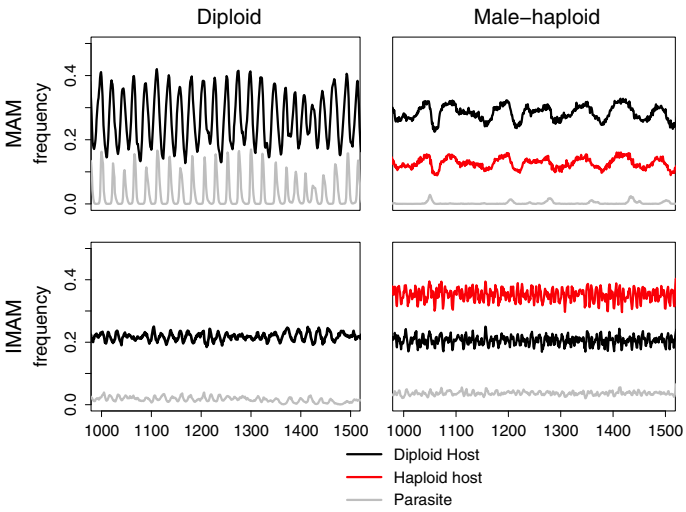


Fig. 2 The change in frequencies of the parasite strain *ab* and the corresponding proportion of the host susceptible to this parasite strain. The host frequencies provided are based on the gender to illustrate the impact of ploidy on the frequency of susceptible hosts within the population. Hence, the observation of a lower frequency of susceptible genotypes under the MAM for the haploid gender and the converse under the IMAM. In these simulations parasite virulence was set to 0.6, transmission rate to 0.7, generation ratio to 10, connectedness to 5 and cliqueness to 0.6. These parameter values were chosen to ensure the potential for high selection pressure on the host population

Under the IMAM the difference in parasite load between the diploid females and haploid males is far more pronounced, leading to large differences in fitness ($w_F = 0.7208$, compared to $w_M = 0.5513$). In contrast under the MAM the overall lower frequency of infected individuals leads to minor overall differences in fitness ($w_F = 0.9965$, compared to $w_M = 0.9985$, against the masking hypothesis). The diploid populations, unsurprisingly, show no difference in fitness between the genders under both interaction models (Kruskal–wallis tests, $p > 0.8$).

The relationships between the transmission parameters and host fitness are all consistently negative. The host-parasite generation ratio tends to have a linear relationship with regards to host fitness (Fig. 4). Whereas the slope appears independent of the combinations between host and interaction model, the same cannot be said of the intercept. This relationship can also be applied to the I–S infection rate, though an asymptote might be reached at intermediate rates (supplementary material, Figure s1). In contrast, the connectedness (or edge density) has a distinctly non-linear relationship when I–S edges are abundant (supplementary material, Figure s2).

Interactions between the model parameters are less evident (Figs. 5, 6). In general host fitness is highest when the parasite-host generation ratios are low and virulence is high (Fig. 5). Small increases in either of these parameters lead to large decreases in host fitness. The lower host fitnesses ranged over a large region of parameter space. No clear evidence of compensation for higher virulence was observed with increases in the host-parasite generation ratio. This is in contrast to the interactions between the different parasite transmission parameters, where compensation appears to be present (Fig. 6).

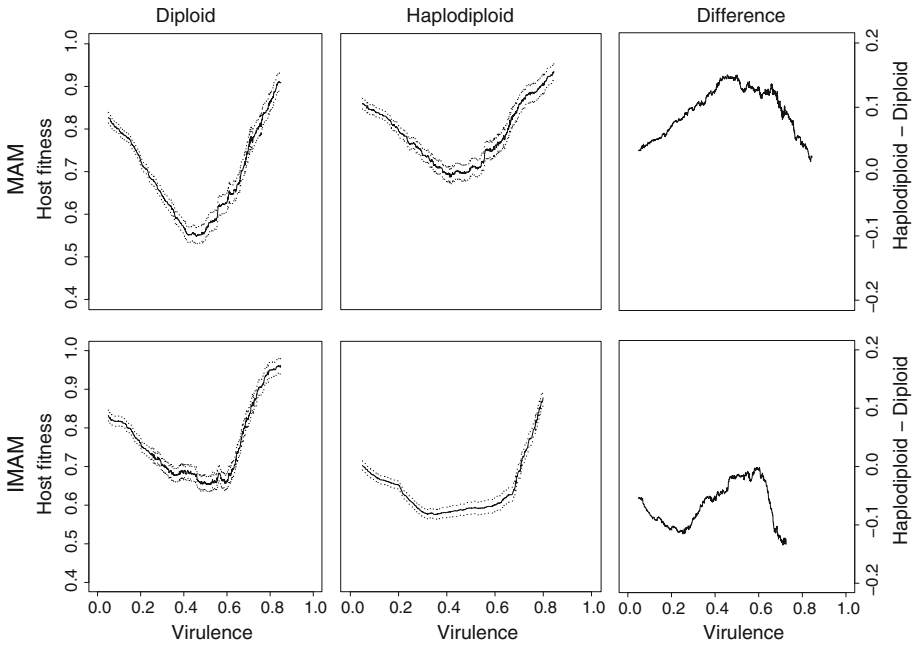


Fig. 3 The relationship between parasite virulence and host fitness. Fitness is calculated systematically throughout the simulation up to 200 times (except when the simulations stopped prematurely from the extinction of multiple host haplotypes). The median and the 95 % confidence interval (*dotted lines*) are plotted, the median and CIs are calculated using a sliding window (± 0.1). The minimum fitness of the diploid population was on average lower than the haploid under the MAM (upper graphs, Wilcoxon V = 224305, p value = 0.000451, median = -0.031); the converse being found under the IMAM (lower graphs, Wilcoxon V = 284909, p value = 0.00015, median = 0.042)

Recombination

In general only small changes in the recombination rate were observed (from 0.5 to either 0.45 or 0.42 in diploid and haplodiploids respectively). However, the distribution of recombination rates is heavily skewed in favour of marginal changes in rate (<0.05, 34.3 % of the simulations). Of those simulations where large decreases in the recombination rate were observed (> 0.1), the majority occurred in haplodiploid host populations (5.1 % compared to 1.5 % in diploids). The observed average rates (including when recombination was driven to <0.1) amoung these simulations were higher in haplodiploid hosts (0.38 under both MAM and IMAM, compared to 0.34 and 0.37 respectively).

GLMM model simplification was used in order to understand which factors are important for understanding the evolution of recombination (within the simulation). The host and interaction model were fitted as random factors, the remaining parameters were treated as fixed with first-order interactions included. After removal of all non-significant terms the model retained: virulence, host-parasite generation ratio, I–S infection rate, and connectedness. Of those terms including cliqueness only the interaction term with parasite virulence was retained (all other terms were retained, see supplementary material s3).

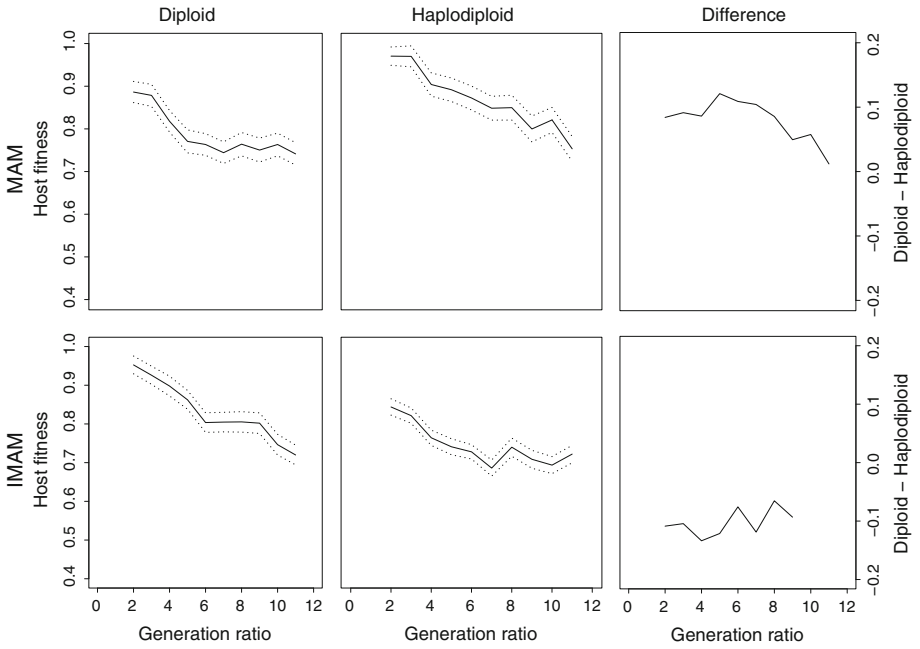


Fig. 4 The relationship of the host—parasite generation ratio and host population fitness. Fitness is calculated as in the previous figure using a sliding window, the 95 % confidence interval is shown by the *dotted lines*. The lower fitness observed for diploids under the MAM (and haplodiploids under the IMAM) is similar to that found for the virulence figure. Here, though, the relationship is strictly negative with a gradual reduction in the degree of divergence between fitness under the different interaction models

Discussion

Interactions between RQ model and ploidy

Host ploidy affects host-parasite evolution directly through the probability of having (MAM), or not having an allele (IMAM) matching those of the parasite. The effect is a change in the frequency of susceptible genotypes: Diploid hosts have lower susceptible frequencies under the IMAM than haploids, with the converse true under the MAM (Fig. 2). These differences in susceptible frequencies lead to profound impacts on host-parasite interactions: parasites fare better under the MAM when hosts are diploid and worse under the IMAM. These results are expected to remain under alternative host-parasite interaction models.

Fitness

When the parasite population fares well, there is a fitness cost to the host population. This is related to three factors: the overall number of infected hosts, the prevalence of multiple infections (Jäger and Schjorring 2006; Popp et al. 2012; Schjorring and Koella 2003; Shykoff and Schmid-Hempel 1991), and the parasite virulence (Fig. 5); as previously reported by Roode et al. (2008): Diploid hosts have lower fitness values over a range of

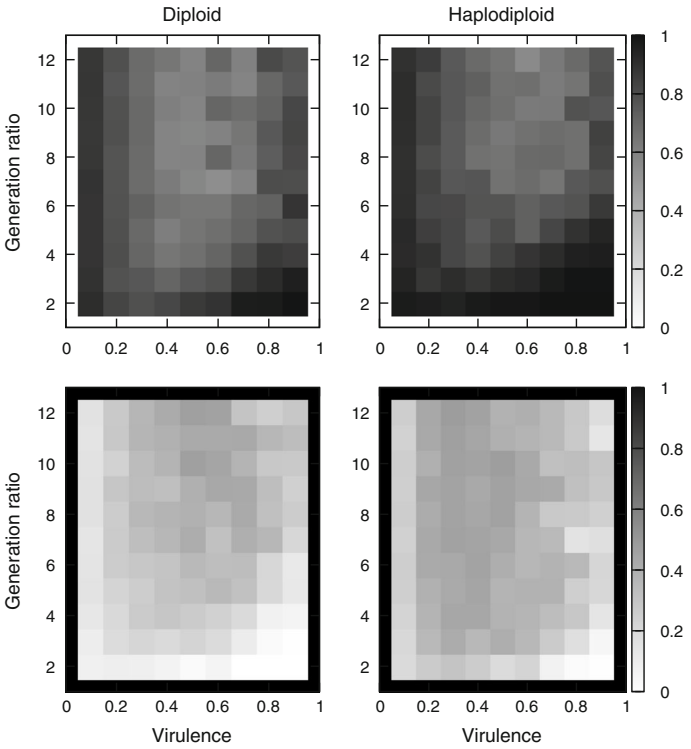


Fig. 5 A heatmap showing the distribution of host fitness depending on the host reproductive system (Diploid, or Haplodiploid) and interaction model (MAM, above; IMAM below). The heatmaps are plotted over a range of transmission values (host-parasite generation ratio) and virulence (probability of death as a result of being infected). NA values are treated as zero: under high transmission rates multiple host genotypes went extinct, in which case the simulation was terminated (producing NAs for fitness). While fitness tends to be some-what higher with lower virulence rates a stronger effect is observable for low transmission rates

parasite virulence under the MAM, higher fitnesses under the IMAM (Nuismer and Otto 2004; Oswald and Nuismer 2007; M'Gonigle and Otto 2011). The converse being true for haplodiploid host populations (higher fitnesses under the MAM than the IMAM). These results being the product of differences in the frequency of susceptible individuals within the host population, in addition to virulence.

The results also show the two-fold influence of parasite virulence on host fitness (Fig. 3). The individual level effect of parasite virulence is through host mortality (or reduced fecundity) reducing the population fitness. The population level effect is through the reduction in parasite transmission through increased host mortality. The interaction between these two effects lead to the expectation of a v-shaped response in host fitness to virulence. The property of host ploidy is expected to change the fitness minima (bottom of the v) in relation to parasite virulence, through changing I–S edge frequencies. More I–S edges would be expected to lead to a fitness minima at higher virulence levels, as transmission becomes less limiting. The results however demonstrate the opposite, the virulence at which the minima occurs is higher under the MAM for diploids (Wilcoxon test,

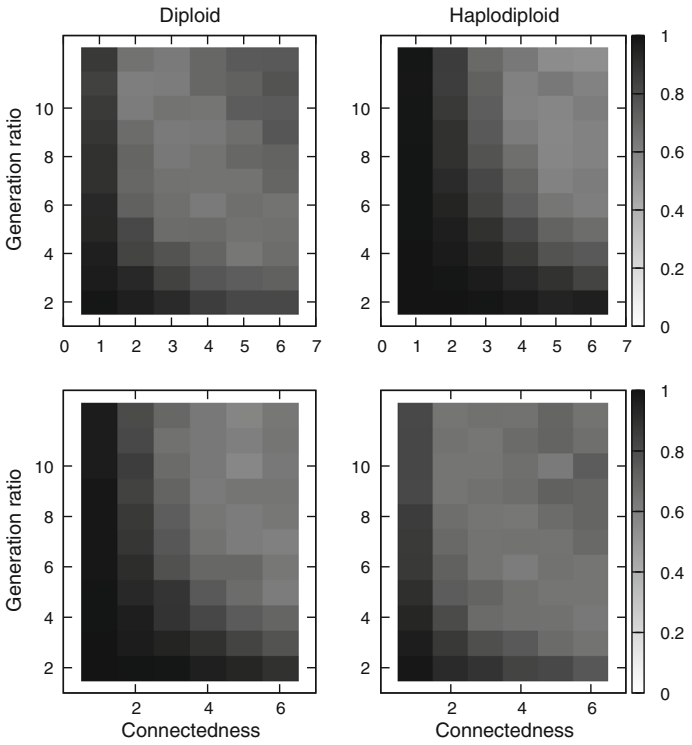


Fig. 6 A comparison using heatmaps of the host fitness depending on reproductive system and interaction model (MAM, above; IMAM, below). The x-axis here uses the average number of connections simulated between host individuals. A strong interaction between the transmission parameters is clear from the results with low generation ratios and connectedness leading to higher host fitness, particularly for haplodiploids under the MAM (*top right*) and diploids under the IMAM (*bottom left*)

W = 540205, p value = 0.0018). In haplodiploid populations the value of parasite virulence associated with the lowest host fitness was higher under the IMAM (Wilcoxon test, W = 457231, p value = 0.0009). A potential explanation is that multiple infections are an important factor when determining population fitness: Multiple infections are more likely for diploids under the MAM and in haplodiploids under the IMAM. If this observed effect is due to multiple infections then it is expected that they are: (a) more important for determining population fitness than single infections, and (b) more sensitive to changing transmission (as a result of increasing virulence) than single infections.

Analysis of fitness over multiple parameters does not change the interpretation of host fitness. This remains lowest over a range of intermediate values for parasite virulence (Fig. 5). A result similar to those from previous studies (King et al. 2012), and in support of the invasion scenarios of M'Gonigle and Otto (2011). The results here being attributable to the proportion of I–S edges, and the general density of edges in the network. In Fig. 6 the effect of edge density can be observed, here this operates in tandem with the host-parasite generation ratio. When either edge density is <2, or the generation ratio <1 host fitness approaches unity. This effect being much weaker in those conditions when the proportion of I–S edges is expected to be large (diploids under MAM, haplodiploids under IMAM).

Recombination

The results presented here are qualitatively in general agreement with previous RQH studies (Agrawal and Otto 2006; Nuismer and Otto 2004; Otto and Nuismer 2004; Peters and Lively 1999, 2007; Schmid-Hempel and Jokela 2002). However, significant quantitative differences are present. Firstly, the frequency with which the direction of selection changes is much lower than that hypothesised to be required for selection on recombination (Barton 1995; Salathé et al. 2007). Secondly, high transmission rates are generally required for selection on recombination (Figures s3 & s4). Thirdly, selection on recombination occurs more frequently in haplodiploid than diploid populations. Lastly, when there was selection acting on the recombination rate the GA predicts higher recombination rates within haplodiploid hosts than diploid. This last result is supported by experimental findings of high recombination rates within haplodiploid species (Wilfert et al. 2007; Meznar et al. 2010; Stolle et al. 2011).

Network versus mean field versus single challenge models

Two general factors are important when considering the type of model to use: (1) model complexity, (2) analytical power. While mean field models are both simpler and more powerful analytically, they are also highly dependent upon system conditions. Typically mean field models perform badly when edge density is low, or when the degree of locality (cliqueness) is high (Keeling 1999). In the current system two factors may suggest a role for network-based models: (1) the significance of the cliqueness by virulence interaction term, (2) the observation of fitness minima at lower values of virulence with more I–S edges. The second observation is also a strong argument against the use of the 'single challenge models'. Mean field and network models inherently include models of parasite distribution, whereas these are absent from the 'single challenge models'.

Acknowledgments Funding for the research was provided by the Deutsche Forschungsgemeinschaft within the priority program SPP 1399 and by yDiv, the Synthesis Centre for Biodiversity Sciences—a unit of the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, also funded by the Deutsche Forschunggemeinschaft (FZT 118).

References

Agrawal A, Lively C (2002) Infection genetics: gene-for-gene versus matching-alleles models and points in between. Evol Ecol Res 4:79–80

Agrawal A, Otto S (2006) Host-parasite coevolution and selection on sex through the effects of segregation. Am Nat 168(5):617–629

Agrawal AF (2009) Differences between selection on sex versus recombination in red queen models with diploid hosts. Evolution 63(8):2131–2141

Ames G, George D, Hampson C, Kanarek A, McBee C, Lockwood D, Achter J, Webb C (2011) Using network properties to predict disease dynamics on human contact networks. Proc R Soc B 278:3544–3550

Barton N (1995) A general model for the evolution of recombination. Genet Res 65:123–144

Boots M, Mealor M (2007) Local interactions select for lower pathogen infectivity. Science 315:1284–1286

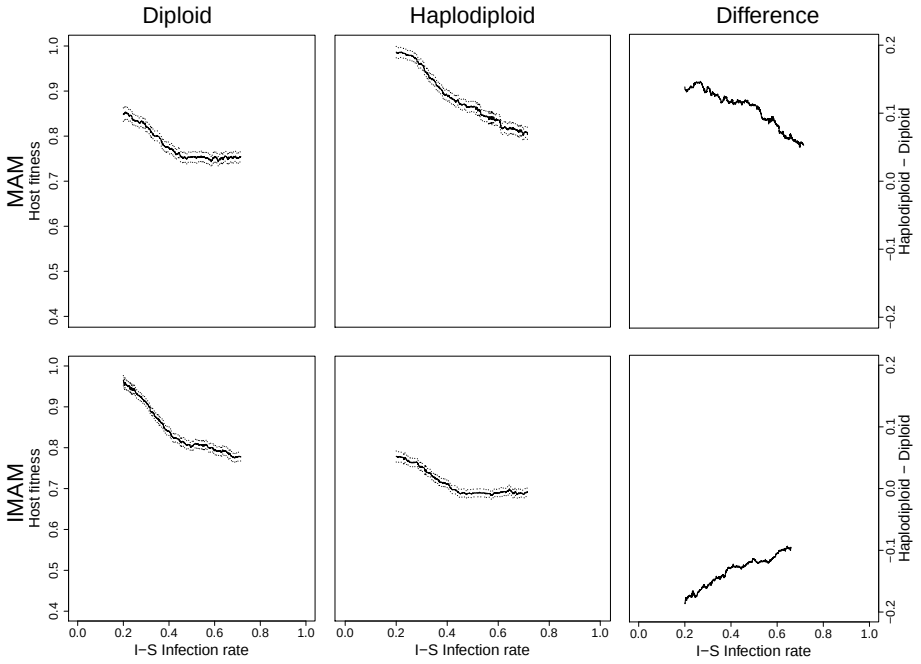
Carroll L (1960) 2 The garden of live flowers. Through the looking-glass and what alice found there (The Annotated Alice: Alice's Adventures in Wonderland and Through the Looking-Glass, illustrated by John Tenniel, with an Introduction and Notes by Martin Gardner ed.). The New American Library, New York

Crawley M (2005) Statistics: an introduction using R. Wily, West Sussex

Engelstädter J, Bonhoeffer S (2009) Red queen dynamics with non-standard fitness interactions. PLOS Comput Biol 5(e1000):469
Hodgeson E, Otto S (2012) The red queen coupled with directional selection favours the evolution of sex. J Evol Biol 25:797–802
Jäger I, Schjorring S (2006) Multiple infections: Relatedness and time between infections affect the establishment and growth of the cestode schistocephalus solidus in its stickleback host. Evolution 60(3):616–622
Keeling M (1999) The effects of local spatial structure on epidemiological invasions. Proc R Soc B 266:859–867
Kidner J, Moritz R (2013) The red queen process does not select for high recombination rates in haplo-diploid hosts. Evol Biol 40(3):377–384
King K, Seppälä O, Neiman M (2012) Is more better? Polyploidy and parasite resistance. Biol Lett 8:598–600
Lively C (2010) An epidemiological model of host-parasite coevolution and sex. J Evol Biol 23:1490–1497
Meznar E, Gadau J, Koeniger N, Rueppell O (2010) Comparative linkage mapping suggests a high recombination rate in all honeybees. J Hered 101:s118–s126
M'Gonigle L, Otto S (2011) Ploidy and the evolution of parasitism. Proc R Soc B 278:2814–2822
Normark BB (2003) The evolution of alternative genetic systems in insects. Annu Rev Entomol 48:397–423
Nuismer S, Otto S (2004) Host-parasite interactions and the evolution of ploidy. PNAS 101:11,036–11,039
Oswald B, Nuismer S (2007) Neopolyploidy and pathogen resistance. Proc R Soc B 274:2393–2397
Otterstatter MC, Thomson JD (2007) Contact networks and transmission of an intestinal pathogen in bumble bee (Bombus impatiens) colonies. Oecologia 154:411–421
Otto S, Marks J (1996) Mating systems and the evolutionary transition between haploidy and diploidy. Biol J Linn Soc 57:197–218
Otto S, Nuismer S (2004) Species interactions and the evolution of sex. Science 304:1018
Peters A, Lively C (1999) The red queen and fluctuating epistasis: a population genetic analysis of antagonistic coevolution. Am Nat 154:393–405
Peters A, Lively C (2007) Short- and long-term benefits and detriments to recombination under antagonistic coevolution. J Evol Biol 20:1206–1217
Popp M, Erler S, Lattorf HMG (2012) Seasonal variability of prevalence and occurrence of multiple infections shape the population structure of crithidia bombi, an intestinal parasite of bumblebees (bombus spp.). MicrobioligyOpen 1(4):362–372
R Core Team (2013) R: a language and environment for statistical computing. R Fundation for Statistical Computing, Vienna, Austria. http://www.R-project.org
Roode JD, Yates A, Altizer S (2008) Virulence-transmission trade-offs and population divergence in virulence in a naturally occuring butterfly parasite. Proc Nat Acad Sci USA 105:7489–7494
Salathé M, Kouyos R, Regoes R, Bonhoeffer S (2007) Rapid parasite adaptation drives selection for high recombination rates. Evolution 62:295–300
Schjorring S, Koella JC (2003) Sub-lethal effects of pathogens can lead to the evolution of lower virulence in multiple infections. Proc R Soc Lond Ser B 270(1511):189–193
Schmid-Hempel P, Jokela J (2002) Socially structured populations and evolution of recombination under antagonistic coevolution. Am Nat 160(3):403–408
Shykoff J, Schmid-Hempel P (1991) Incidence and effects of four parasites in natural populations of bumble bees in switzerland. Apidologie 22:117–125
Stolle E, Wilfert L, Schmid-Hempel R, Schmid-Hempel P, Kube M, Reinhardt R, Moritz R (2011) A second generation genetic map of the bumblebee bombus terrestris (linnaeus, 1758) reveals slow genome and chromosome evolution in the apidae. BMC Genom 12:48
Sutton J, Nakagawa S, Robertson B, Jamieson I (2011) Disentangling the roles of natural selection and genetic drift in shaping variation at mhc immunity genes. Mol Ecol 20:4408–4420
Takahashi Y, Morita S, Yoshimura J, Watanabe M (2011) A geographic cline induced by negative frequency-dependent selection. BMC Evol Biol 11:256–266
Webb S, Keeling M, Boots M (2013) A theoretical study of the role of spatial population structure in the evolution of parasite virulence. Theor Popul Biol 84:36–45
Wilfert L, Gadau J, Schmid-Hempel P (2007) The genetic architecture of immune defense and reproduction in male bombus terrestris bumblebees. Evolution 61:804–815
Williams T, Kelly C (2003) GNUPLOT: an interactive plotting program. http://www.gnuplot.info

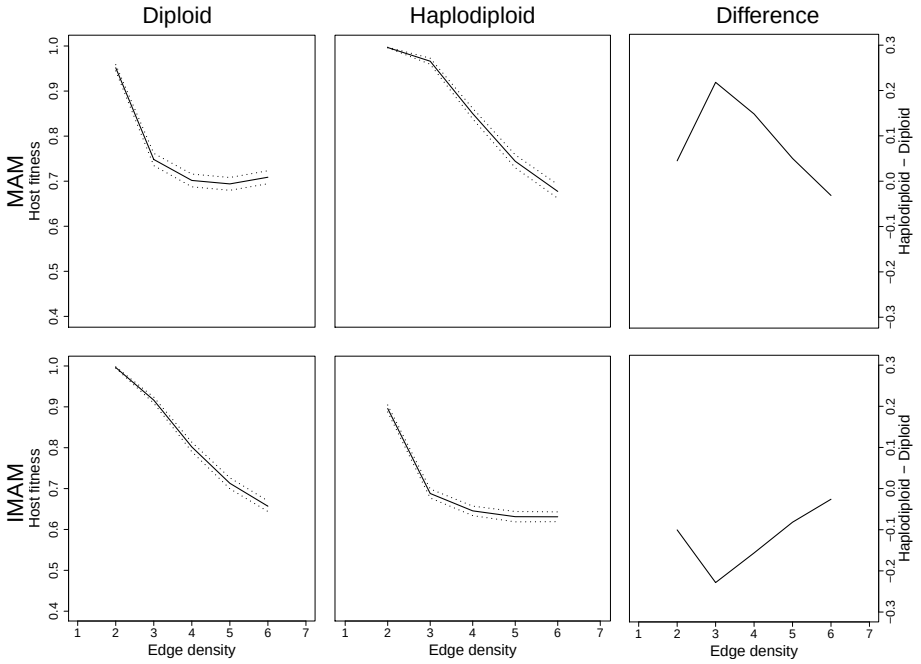
Supplement 1

Figure 1. The results of host fitness (over I – S infection rate) similar to figures 3 & 4 in the paper. The differences in fitness between the diploid and haplodiploid populations are in the same direction as the plots with parasite virulence and generation ratio. Additionally the gradient of the slopes are higher for the haplodiploids under the MAM and diploids under the IMAM (similar to the results with generation ratio).



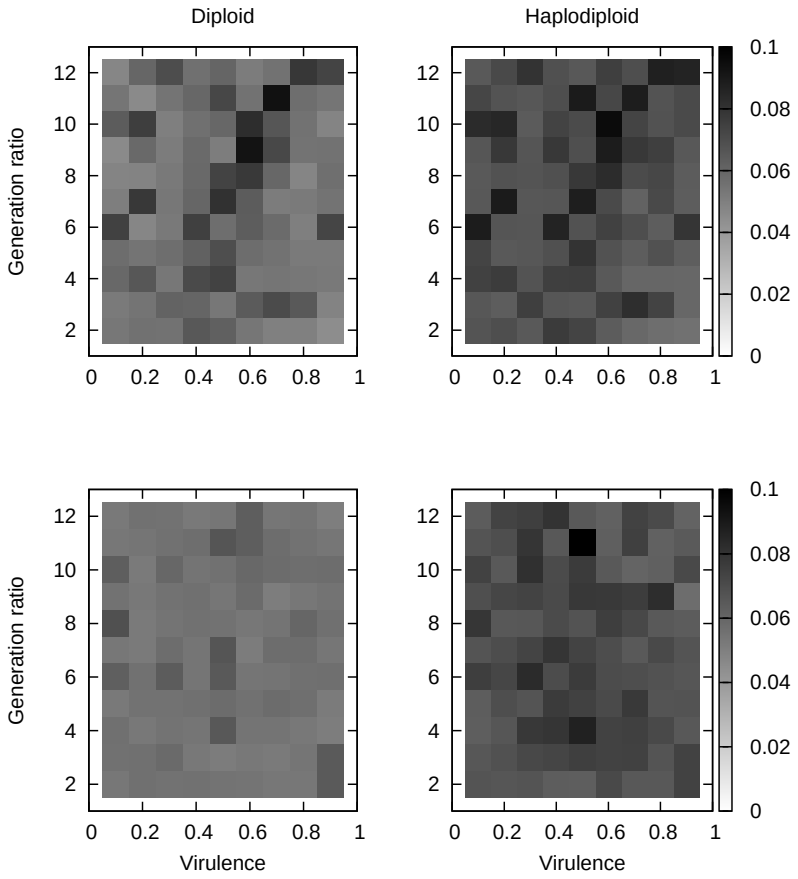
Supplement 2

Figure 1. The change in host fitness over a range of edge densities (Connectedness). This figure is produced similarly to 3 & 4 within the main text and S1. These results show stronger responses with lower values of connectedness when there is a large proportion of I − S edges within the host population. Differences between the host populations reduce, as in the previous plots, with higher levels of host connectedness.



Supplement 3

Figure 1. An illustration of the degree to which the recombination rates decayed in the simulation from free recombination (0.5). Given the normally distributed mutation term of mean 0 and standard deviation 0.01 values greater than 0.04 on the colour axis indicate the presence of selection against the initial recombination rate. While differences are observable between the two host systems and interaction model (MAM above, IMAM below), no distinct pattern appears to be present regarding the parameter combination.



Supplement 4

Figure 1. A corresponding figure for the degree of edge density and parasite transmission rates on the decay of the recombination rates. While no clear patterns for selection could be seen for a combination of virulence and transmission, here a clear dependence on both higher edge density and transmission rates is apparent the decay in recombination rates in haplodiploid hosts. With a diploid host population under the MAM the converse seems to be true, while this relationship is either reversed or not apparent for these hosts under the IMAM.

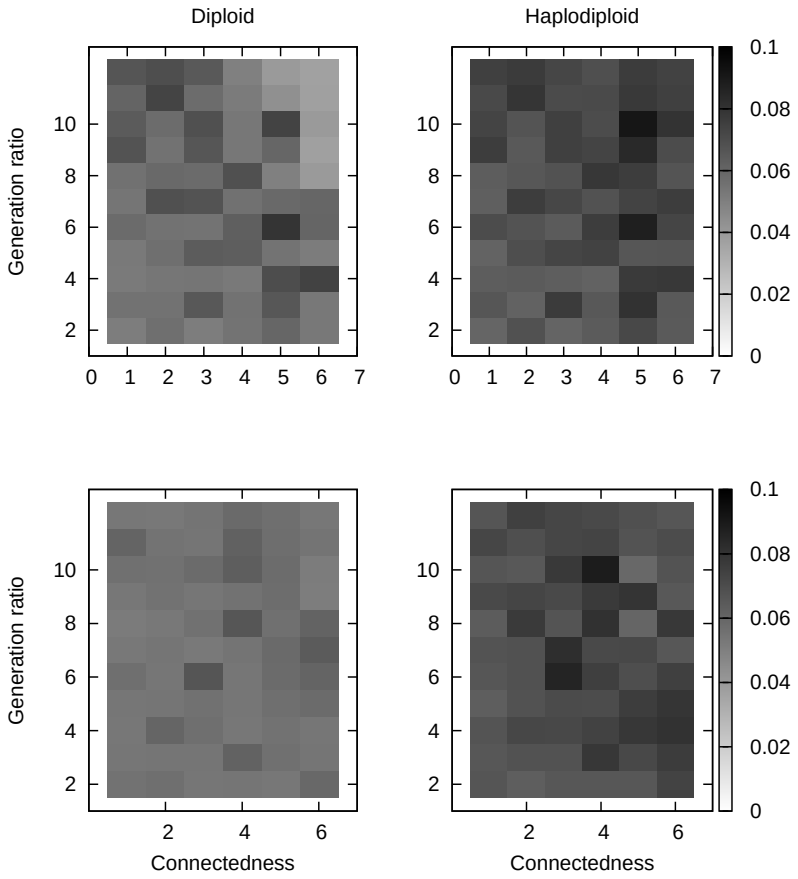


Table 1. Common genetic interaction models. Two of the most commonly used host-parasite interaction mechanisms in the RQ literature. Under the MAM the parasite can infect a host, if the parasite's alleles match the respective host's alleles. The converse is true for the IMAM, akin to the adaptive immune systems of Vertebrates (MhC) where the host needs to recognise the parasite alleles to resist infection. The ability of a parasite to infect the host being indicated by unity (1, host resistance is given by 0) in the following tables.

Haploid-Haploid interactions

	Parasite MAM			
Host	*ab*	*aB*	*Ab*	*AB*
ab	1	0	0	0
aB	0	1	0	0
Ab	0	0	1	0
AB	0	0	0	1

	IMAM			
Host	*ab*	*aB*	*Ab*	*AB*
ab	0	1	1	1
aB	1	0	1	1
Ab	1	1	0	1
AB	1	1	1	0

Diploid hosts - Haploid parasites interactions

	Parasite MAM			
Host	*ab*	*aB*	*Ab*	*AB*
ab/*ab*	1	0	0	0
2*ab*/*aB*	1	1	0	0
2*ab*/*Ab*	1	0	1	0
2*ab*/*AB*	1	1	1	1
aB/*aB*	0	1	0	0
2*aB*/*Ab*	1	1	1	1
2*aB*/*AB*	0	1	0	1
Ab/*Ab*	0	0	1	0
2*Ab*/*AB*	0	0	1	1
AB/*AB*	0	0	0	1

	IMAM			
Host	*ab*	*aB*	*Ab*	*AB*
ab/*ab*	0	1	1	1
2*ab*/*aB*	0	0	1	1
2*ab*/*Ab*	0	1	0	1
2*ab*/*AB*	0	0	0	0
aB/*aB*	1	0	1	1
2*aB*/*Ab*	0	0	0	0
2*aB*/*AB*	1	0	1	0
Ab/*Ab*	1	1	0	1
2*Ab*/*AB*	1	1	0	0
AB/*AB*	1	1	1	0

Table 1. The number of simulations where the recombination rate decayed below a value of 0.4. This value was arbitrarily chosen to pick those cases in which selection was relatively consistent through the simulation. Choosing values to 0.3 shows a consistently higher occurrence of selection on recombination within the haplodiploid host populations, dependent upon the interaction model. These results are significantly divergent between host and interaction model according to a x^2 test ($x^2 = 35.14$, df = 1, p-value = $3.1e^{-9}$)

MAM	IMAM	
Diploid	8 (5 – 18)	13.5 (11 – 16)
Haplodiploid	13 (9 – 19)	10 (4 – 16)

Pseudocode

Simulation. The following pseudocode describes the key algorithms that we used. While we do not describe the structure of the code, through the basic use of these algorithms it would be possible to create an equivalent simulation using the following pseudocode. One of the first points to be made about the following pseudcode is that each block represents a basic step in the calculations. Though a block may represent multiple, or part of a function in the actual code.

Basic pseudocode structure We tend to use some basic methods of separating functions, variables and model parameters within the pseudocode. Any variable or parameter is identified by italicization (new_edge), with multi-word parameters connected by an underline. Generally the parameters present in the pseudocode represent state variables, exceptions to this are: max_edge_no, min_edge_no, $transmission$, $generation\_ratio$, $parasite\_virulence$, Sex_ratio, $genotype$ (both male and female, $genotype_m$ and $genotype_f$), $locus$, and $recombination$. In contrast a function is represented by normal text (starting capitalized) preceding an open and closing bracket (Random()). Comments on a function or parameter are provided inside braces , comments are given to identify the value of a parameter. Assignment and accession are symbolised as is standard for pseudcode (left arrows for assignment, and right arrows for accessing a value, or trait). If we are accessing an item from an array of items then the iterator that we use to identify the position within the array is indicated by sub-script. Super-scripts are used for more specific identification purposes for similar parameters between different objects, the only exception is when we use a parameter to power to which a value is multiplied, such a parameter will be enclosed in brackets. In boolean conditions we use the equals (=) sign to indicate the equivalent to boolean operator, and the standard AND boolean operators. In the last section of the pseudecode we present a new kind of parameter written entirely capitalized, these parameters are all containers (they contain the individual objects that we use in all other segments of the pseudocode).

Network construction (Algorithms 1–3)

Creating a network required more information than was available from the literature. As a result we took one of the better known algorithms for constructing networks (the scale-free network construction method). These methods tend to form networks with a heavily skewed distribution featuring a few super-spreaders (a distribution that contrasts with those observed for social insects[3]), to counter this

we included a diminishing returns term. However, due to the random generation of networks the amount of variance observed within the simulations still remained rather high due to random network topologies (despite refinement of the networks). To counter this we generated multiple random network topologies (refined to the same parameter conditions) during single simulation runs.

Algorithm 1 Connection probability

new_edge a node parameter indicating likelihood of forming a new edge
no_edge a node parameter indicating the number of edges present {these parameters are always set to zero when constructing a new network}
$new\_edge \leftarrow new\_edge + \frac{1}{no\_edge}$;
$no\_edge \leftarrow no\_edge + 1$;

Algorithm 2 Adding edges

node refers to the currently added node
nodes refers to the array of nodes
still_to_add how many edges to add {An input variable}
edges the number of edges so far added {this algorithm is used when adding a new node (except the first 2 nodes)}
while $edges < still\_to\_add$ **do**
  $likelihood \leftarrow$ Random (Sum (new_edge));
  $which \leftarrow 0$;
  $where \leftarrow 0$;
  for all Present nodes **do**
    if $(which \leftarrow which + node \rightarrow new\_edge_{where}) < likelihood$ **then**
      $node \rightarrow$ Add edge($nodes_{where}$);
      $nodes_{where} \rightarrow$ Add edge($node$);
    end if
    $where \leftarrow where + 1$;
  end for
end while

Algorithm 3 Adding nodes

current_node is a class variable indicating the position of the current node
input_variables are a set of variables for the biological traits of each node {The nodes technically exist prior to the execution of this algorithm, hence we copy information to them}
$no\_edges \leftarrow 0$;
Set($current\_node, input\_variables$);
$no\_edges \leftarrow min\_edge\_no +$ Random($max\_edge\_no - min\_edge\_no$);
Add edges(no_edges);

Parasite transmission (Algorithms 4–6)

Algorithm 4 Initial population

hosts is a class variable that contains the addresses of host objects
previous_count has the value of the previous number of hosts infected
no_susceptible is the current number of susceptible hosts
proportion_susceptible is the proportion of the susceptible host population
Statistics is a variable containing the parasite strain statistics
iterators is an iterator to go through the susceptible part of the host population {the *iterator* variable is randomized}
$infect \leftarrow previous\_count \times proportion\_susceptible$;
if $infect > no\_susceptible$ **then**
  $infect \leftarrow no\_susceptible$;
  Statistics → Update(*no_susceptible*);
else if $infect = 0$ **then**
  $infect \leftarrow 1$;
  Statistics → Update(1);
end if
$i \leftarrow 0$;
while $i < infect$ **do**
  $hosts_{iterators_i}$ → Infect(this);
  $i \leftarrow i + 1$;
end while

Algorithm 5 Transmission between hosts

self is the host object under consideration
strain one of 4 types of parasite strain
neighbours contains an array of all the connecting nodes {'*neighbour*' is a singular node from *neighbours*}
no_neighbours states how many connecting nodes exist
if *self* → Is Infected() *AND self* → Is Alive() **then**
  for all Parasite strains **do**
    if *self* → Is Infected(*strain*) **then**
      'Is Infected()' takes two arguments, identifying either general infection, or infection by a specific strain
      neighbours ← *self* → Neighbours();
      no_neighbours ← *self* → No. Neighbours();
      for all *neighbours* **do**
        probability ← Random(1);
        if *neighbour* → Alive() *AND neighbour* → Compatible(*strain*) *AND probability* < *transmission* **then**
          neighbour → Infect(*strain*);
        end if
      end for
    end if
  end for
end if

We treated parasite transmission processes to be independent between the strains. This treatment aided in the coding by eliminating the impact of order dependence within the simulation, otherwise we

the method of parasite transmission used is standard for epidemiological models [1].

Algorithm 6 Parasite generations

generation is a variable whose value is the parasite generation number
generation_ratio is a variable describing the host-parasite generation ratio
$Generations \leftarrow \frac{1}{generation\_ratio}$;
hosts is a variable that contains addresses to all of the nodes in the network
while $generation < Generations$ **do**
  for all *hosts* **do**
    Execute the transmission algorithm
  end for
end while

Host reproduction

Generally with network models host reproduction is modeled as the regeneration of current networks. However, as regarded the high variation observed from the network topologies and the subsequent multiple generation of network topologies for the separate host generations, we dropped involvement of network topology for host reproduction. While this will have a strong impact on the host genetic composition and between host generation, the intra-generation transmission still keeps to a range defined by the network topology parameters.

Host death (Algorithms 7–8)

The first step that we take on host reproduction is to calculate which hosts survive to contribute to the next generation, this step also being linked to identifying the between host generation transmission rates. As parasite transmission was considered independent between parasite strains so was parasite virulence, the chance of dying multiplying with the product of virulence rates for every recorded infection, Algorithm 7. If hosts died then they were dropped from the host containers and could not be randomly selected as parents for the next generation

Generating offspring (Algorithms 9–11)

In order to control for fluctuations occuring in the effective population sizes between males and females (N_{e^m} and N_{e^f}), the selection of individuals was randomised from similarly sized populations (Algorithm 9). The subsequent selection process (of those individuals allready selected to contribute) however used

Algorithm 7 Mortality risk

This algorithm is operated for each host node upon host reproductive event
$infected$ indicates whether a host is infected, or not
alive indicates whether the host is alive, or not
parasite_virulence is the proportion of hosts killed by a single strain {*parasite_virulence* is assumed to have a cumulative effect}
parasite_loss is a handler to an array of losses to the parasite population
$infections$ value is an array indicating infection by strains i,..n
strain is a variable indicating the strain under consideration
no_strains is a variable whose value is the number of strains
$survival \leftarrow 1$; {*survival* is a variable indicating the expected probability of survival}
if $infected\ AND\ alive$ **then**
  $strain \leftarrow 1$;
  while $strain < no\_strains$ **do**
    if $infections_{strain}$ =TRUE **then**
      $survival \leftarrow survival \times (1 - parasite\_virulence)$;
      $parasite\_loss_{strain} \leftarrow 1$;
    else
      $parasite\_loss_{strain} \leftarrow 0$;
    end if
    $strain \leftarrow strain + 1$;
  end while
end if
$probability \leftarrow$ Random(1); {Choose a random real number from 0 to 1}
if $probability > survival$ **then**
  return 0;
end if
return 1;

Algorithm 8 Calculating death rates

hosts is a list of the nodes in the network
parental is a container class that identifies reproductively capable hosts
parasite_loss is a handler to an array that contains the parasite strains lost {*parasite_death* is similar but contains the results for individual hosts}
for all *hosts* **do**
  if Mortality risk($host, parasite\_death$) **then**
    $parental \rightarrow$ Add element($host$);
  else
    for all *strains* **do**
      $parasite\_loss_{strain} \leftarrow parasite\_loss_{strain} + parental \rightarrow$ Is strain present($strain$); {Increments the value of *parasite_loss* by one if the strain is present}
    end for
  end if
end for

replacement. From the selected individuals recombinant fractions were chosen randomly according to the pseudocode of algorithm 9, a system that works for the two locus system employed here.

Algorithm 9 Recombinant fractions

Genotypes is a prior constructed structure containing recombinant genotypes
Recombinant_events the number of such events for each *Genotype*
$Non - recombinant\_events$ the number of such events for each *Genotype*
no_Genotypes the number of different haplotypes
probability ← Random(2.0);
possibilities ← 0;
recombination the recombination rate of the current entity {Holds the cumulative value of the probabilities}
if *no_Genotypes* > 1 **then**
  for all *Genotypes* **do**
    Recombinant_events ← *Genotype* → No. events();
    $Non - recombinant\_events \leftarrow loci - Recombinant\_events - 1$;
    $possibilities \leftarrow possibilities + (1 - recombination)^{Non-recombinant\_events} * (recombination)^{Recombinant\_events}$;
    if *probability* < *possibilities* **then**
      return *Genotype*;
    end if
  end for
end if
return $Genotype_0$;

Algorithm 10 Combining gametes

paternal the address of the paternal entity
$genotype_p$ the genotype of the paternal gamete (recombinant, or non-recombinant)
maternal address of the maternal entity
$genotype_m$ genotype of the maternal gamete (recombinant, or non-recombinant) {Unless the above variables are referred to, all subsequent variables are from the offspring}
recombination ← (*paternal* → *recombination* + *maternal* → *recombination*)/2 + Mutate(); {*recombination* is kept within the bounds; 0, 0.5. Mutate() generates variance around a mean of 0}
for all *maternal* → *loci* **do**
  $genotype_{locus} \leftarrow genotype^p_{locus}$;
  $genotype_{locus+loci} \leftarrow genotype^m_{locus}$;
end for{Creating a haploid offspring is similar except that all paternal elements are substituted by the maternal element}

model execution (Algorithm 12)

The core of the simulation involves in order, creating the new generation of hosts, constructing the network, simulating parasite transmission and finally host death and reproduction. In this frame work two containers were used for storing the host information (which were switched on generating each new host generation) enabling comparisons between the previous and proceeding generations.

Algorithm 11 Collecting and randomizing hosts

$iterator^h \leftarrow$ Calculate death rates(); {The iterator arrays are generated within Alg. 8 (by the container class '*parental*')}
$male\_iterators \leftarrow iterator^h \rightarrow$ Get(M); {*male-* and *female-* $\_iterators$ are iterator arrays for sampling parental entities}
$sample\_size_{male} \leftarrow male\_iterators \rightarrow$ Size(); {$sample\_size$ of both types indicate the length of the iterator arrays}
$female\_iterators \leftarrow iterator^h \rightarrow$ Get(F);
$sample\_size_{female} \leftarrow sample\_size_{male}/Sex\_ratio$;
if $female\_iterators \rightarrow$ Size() $< sample\_size_{female}$ **then**
  $sample\_size_{female} \leftarrow female\_iterators \rightarrow$ Size();
  $sample\_size_{male} \leftarrow female\_iterators \times Sex\_ratio$;
end if
$male\_iterators \rightarrow$ Shuffle(); {'Shuffle' randomizes the order of elements one at a time}
$female\_iterators \rightarrow$ Shuffle();
while Generating offspring **do**
  $collect\_father \leftarrow$ Random($sample\_size_{male}$); {Choose a random male individual}
  $collect\_mother \leftarrow$ Random($sample\_size_{female}$); {'$collect\_father$' and '$collect\_mother$' are indices to be used for collecting positions from the iterator arrays}
  Get recombinant fractions using Recombinant fractions Algorithm and the $collect\_mother$ and $collect\_father$ indices;
  Use Combining gametes algorithm to combine the gametes from the Recombinant fractions algorithm;
end while

$PREVIOUS\_GENERATION \rightarrow$ Clear(); {'Clear' removes the information but not the data of the calling object}
$HOST\_NETWORK \rightarrow$ Reset(); {Similar to 'Clear', except information is not overwritten/removed}
$CURRENT\_GENERATION \rightarrow$ Randomize(); {$CURRENT\_GENERATION$ contains the actual entities created by the algorithms under "Generating offspring"}
for all $CURRENT\_GENERATION \rightarrow HOSTS$ **do**
  $HOST\_NETWORK \rightarrow$ Insert($CURRENT\_GENERATION \rightarrow HOST$); {Each host in $HOSTS$ is inserted into the biological network}
end for
$HOST\_NETWORK \rightarrow$ Refine(); {Refines the network to the correct topology}
for all Parasite generations **do**
  Execute the parasite algorithms
end for
for all $CURRENT\_GENERATION \rightarrow HOSTS$ **do**
  Execute host death algorithms
end for
$CURRENT\_GENERATION \leftarrow$ Switch generation($CURRENT\_GENERATION$); {$CURRENT\_GENERATION$ is switched with $PREVIOUS\_GENERATION$}
for all $CURRENT\_GENERATION \rightarrow HOSTS$ **do**
  Execute host reproduction algorithms
end for

References

[1] M.J. Keeling. The effects of local spatial structure on epidemiological invasions. *Proceedings of the Royal Society B*, 266:859–867, 1999.

[2] J. Kidner and R.F.A. Moritz. The red queen process does not select for high recombination rates in haplodiploid hosts. *Evolutionary Biology*, 40(3):377–384, 2013.

[3] Michael C. Ottersatter and James D. Thomson. Contact networks and transmission of an intestinal pathogen in bumble bee (*bombus impatiens*) colonies. *Oecologia*, 154:411–421, 2007.