The Evolution of haplo-diploidy and its consequences for the Red Queen Hypothesis (RQH).

Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.)
der

Naturwissenschaftlichen Fakultät I - Biowissenschaften -

der Martin-Luther-Universität Halle-Wittenberg,

vorgelegt

von Herrn Jonathan Kidner

geb. am 11.04.1984 in Gloucester

Abstract:

Haplo-diploidy can be found in the Insects, Arachnida and Rotiera; having evolved convergently at least 20 times throughout these groups, involving multiple mechanisms (Normark 2003). Subsequently the research involved in this thesis deals with simulation and model based investigations of the evolution/invasion of haplo-diploidy, in addition to a more lengthy into its impacts on host-parasite coevolution. And the consequences of haplo-diploidy on the red queen hypothesis (the evolution of sex and recombination in co-evolving species). Demonstrating a strong qualitative and quantitative impact of haplo-diploidy on the red queen hypothesis. Therefore, I argue the case for greater evolutionary interest in and of haplo-diploidy, and alternative genetic systems.

Abstrakt:

Haplodiploidie tritt bei Insekten, Spinnentieren und Rädertierchen auf. Sie hat sich innerhalb dieser gruppen mindestens 20 mal unabhängig voneinander und durch unterschiedlichste Mechanismen entwickelt (Normark 2003). Die vorliegende Arbeit beschäftigt such mit simulations- und modelbasierten Untersuchungen zur Invasion/Evolution der Haplodiploidie, in Ergänzung zu einer ausführlicheren Betrachtung ihrer Auswirkungen auf die Koevolution von Parasit und Wirtsorganismus, sowie sich daraus ergebenden Konsequenzen für die Red-Queen-Hypothese (Hypothese zur Evolution der sexuellen Fortpflanzung und Rekombination in co-evolvierenden Spezies). Es offenbarten sich starke qualitative und quantitative Effekte von Haplodiploidie auf die Red-Queen-Hypothese. Aus diesem Grund spreche ich mich für ein stärkeres evolutionsbiologisches Interesse an Haplodiploidie im Speziellen und alternativen genetischen Systemen im Generellen aus.

Keywords:

Invasion model, evolution, haplo-diploidy, male-haploidy, red queen hypothesis, sex, recombination.

Keyworten:

Invasion Model, Evolution, Haplo-diploidie, Male-Haploidie, rote Könige Hypothese, Sex, Rekombination

Referees:

1. Prof. Dr. Tanja Schwander, Department of Ecology and Evoltion, UNIL Sorge, Le Biophore, CH - 1015, Lausanne, Switzerland

2. Prof. Dr. Ido Pen, GELIFES - Groningen Institute for Evolutionary Life Sciences Nijenborgh 7, 9747, AG Groningen, the Netherlands

3. Prof. Dr. Robin FA Moritz, Fachbereich Biology, Institute of Biology, Martin-Luther-University, Halle/S., Sachsen-Anhalt, Germany

Jonathan Henry Kidner

Personal details: **Date of birth:**

Contact details: **Address:** 23 Streiberstraße
Halle/Saale
Postcode: 06110
Mobile telephone: +49 0176 38049072
Email: jonathan.kidner@zoologie.uni-halle.de

Education:
University: Martin Luther Universitat, Halle-Wittenberg
Qualification: PhD in Biological Sciences
Chairman of the Committee: Prof. Dr. J. Krieger
Date of defence: 19.12.2016
University: University of York
Start date: 2008
Completion date: 2009
Qualification: Master by research in computational biology

University: Edinburgh University
Start date: 2006
Completion date: 2007
Qualification: Master of science in quantitative genetics and genome analysis

University: University of Reading
Start date: 2002
Completion date: 2005
Qualification: Bachelor of science in evironmental biology

SIGNATURE: DATE:

Contribution to the doctoral dissertation — I can declare that I am the sole contributor to the content contained within the dissertation and have adequately given credit where it is due (for data sources used within).

Contents

List of Figures

Part I

Introduction

General thesis outline

This thesis represents the sum of three theoretical studies on the evolution of male-haploidy and the consequences of male-haploidy on host pathogen co-evolution in the Arthropods. I first investigate two of the major hypotheses that have been proposed for the origin and maintenance of haplo-diploidy in nature (in part II). Specifically I investigate the importance of *a*) the maternal transmission hypothesis (MTH) (Smith, 2000) and *b*) the deleterious mutation hypothesis (DMH) (Goldstein, 1994), for the origin and maintenance of haplo-diploidy; using an updated model system that incorporates stochastic noise through genetic drift (Chapter II). I subsequently investigate the dynamics of host-parasite interactions in haplo-diploid species, making a direct comparison to diplo-diploid species. In this section both a simple population genetic approach (Chapter II) and a novel network approach (Chapter III) are utilized. In a final part (IV) I discuss the overall conclusions of the studies presented in this thesis, their potential implications and possible future avenues of research.

Haplo-diploidy, or male-haploidy:

Experimental observations and its phylogenetic distribution

Haplo-diploidy as a natural phenomenon is wide-spread. A generalised definition includes the presence of haploid and diploid individuals within a single population at any given time, with the distribution of ploidy (whether an individual is haploid, or diploid) being dependent on its gender. This latter point is important as it separates a large group of plants and fungi, individuals of whom feature separate haploid and diploid life-phases (Mable & Otto, 1998), from the Arthropods. The studies presented here focus on the haplo-diploid sex determination system found in the Arthropods. Although males predominantly represent the haploid gender, female-haploidy is observable in a few instances (Pijnacker et al., 1980; Beukeboom et al., 2007). Though these instances of female-haploidy occur under different sex determination systems that do not interest us here (Pijnacker et al., 1980).

Male-haploidy, or haplo-diploidy is remarkably common in the Arthropoda (Breeuwer, 1997; Normark, 2003; Tortajada et al., 2009): it can be found in spider-mites, rotifers, and multiple insect groups. All in all, it is thought to have independently evolved at least twenty times in the Arthropoda (Borgia, 1980; Normark, 2003), but very likely more often considering the current state of knowledge of many Arthropod lineages. Part of this distribution is illustrated in the phylogenetic study of Misof et al. (2014) on the class Insecta, where the majority of known cases of male-haploidy are observed (Figure 1). The Hymenoptera and Thysanoptera represent very deep rooted male-haploid lineages; Deep evolutionary events, such as these, can be hard to analyse as their signals are likely to be obscured by other events/processes. In other cases the evolution of male-haploidy may not be as deep, but suffer complications from other factors. In the Aleyrodidae (Hemiptera, Figure 1) a huge diversity of sex-systems exist with frequent transitions.The same occurs in the Aphidoidea (Campbell et al., 1994). While

this can be advantageous (increasing the sample size, and reducing the number of correlates) difficulties exist with mapping phenotypic traits to phylogenetic trees: A common issue being the divergence of evolutionary histories between the phenotypic and phylogenetic trees. While divergences between these trees are informative, complex traits (such as male-haploidy) will always be difficult to map to a tree. The simplest technique to map a complex trait would be to assume independence between the traits/data available, however this is unlikely to ever be the case. Dropping independence will result in a massive reduction in the available degrees of freedom.

These features make haplo-diploidy a topic that is hard to cover from an experimental perspective. Therefore I turn to theoretical, computational approaches to investigate its origin and maintenance.

What does haplo-diploidy mean?

It is in consideration of the large diversity of different sex-systems that the first major issue of my research topic is encountered: how to define a haplo-diploid system? This issue, while being primarily limited to resolving the issue of the number of chromosome copies under natural selection, also includes inheritance patterns and sex determination.

To illustrate the difficulty and diversity regarding haplo-diploidy, let us consider the scale-insects. The scale insects include the haplo-diploid lineages: the Iceryini, Aleyrodidae (which are here placed close to their sister taxon in Fig. 1 (Gullan & Cook, 2007; Misof et al., 2014)) and a large portion of the Neococcoidea, the ancestral lineages of which are known to be diplo-diploid. Within these groups, at least four different mechanisms generating male-haploidy are known: *a*) Arrhenotoky, in which unfertilized eggs develop in to haploid males and fertilized eggs develop as females; *b*) Paternal Genome Elimination (PGE), a composite of three distinct processes - often linked to maternally transmitted endosymbionts - whereby the paternally inherited chromosome set is silenced and gradually lost; *c*) Cytoplasmic Incompatibility (CI) , male-killing bacteria/fungi, both of these alter the sex ratio improving the competitive advantage of male-haploidy (Engelstaedter & Hurst, 2006); *d*) Genomic imprinting, a mechanism that can allow for competition between genders as the genome retains a 'memory of its parental origin (Ross et al., 2010). Furthermore the genetic systems of the scale-insects show extremely high lability, both within PGE and to alternative systems: arrhenotoky, deuterotoky (The parthenogenetic reproduction of both males and females), diploidy and thelytoky (reviewed in Gavrilov (2007)). The causal relationships are, however, difficult to elucidate due to the strong coupling of the genetic systems to their potential drivers (Ross et al., 2010).

In other groups, yet further deviations may be found. Within the Sciarid flies (Diptera, Sciaridae) PGE occurs in germ-line cells, whereas somatic cells retain an almost complete set of paternal chromosomes (Goday & Esteban, 2001). To facilitate analysis I adopt two definitions for PGE: *a*) Germline PGE, similar to the PGE mechanism in Sciarids, and *b*) embryonic PGE, in which the PGE process occurs early in embryo development, rendering male somatic

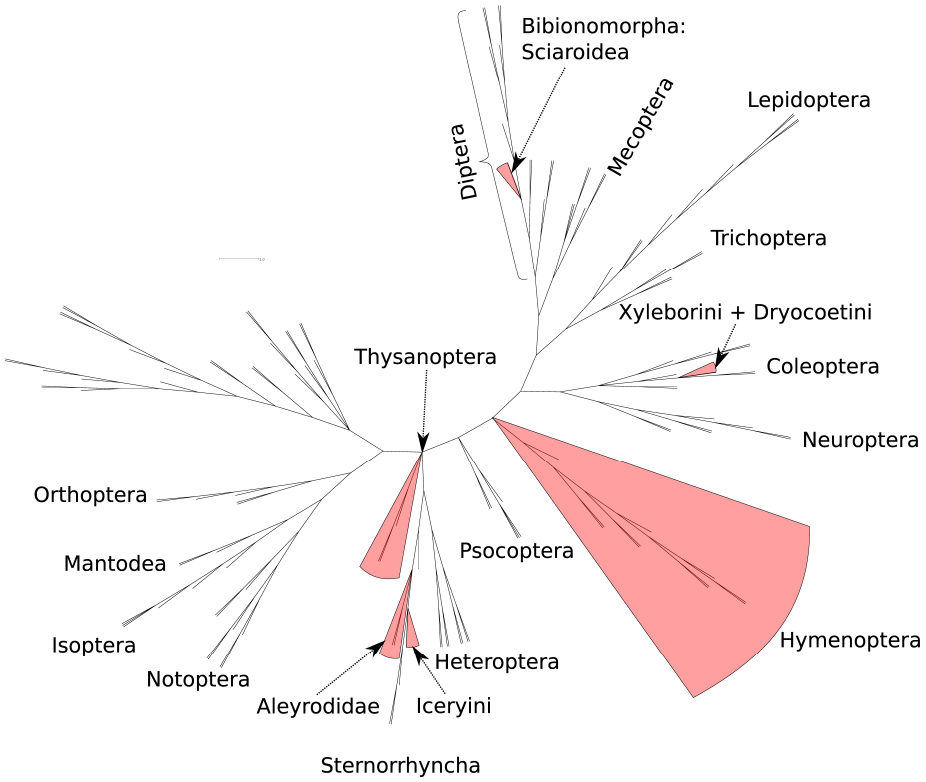

Figure 1: **A phylogenetic tree taken from (Misof et al., 2014) and adjusted to partially demonstrate the distribution of male-haploidy in Insects**. In some instances (Aleyrodidae, Iceryini, Xyleborini and the Sciaroidea) the placements are based on closely related taxa rather than members of these groups. Only a subset of these groups are shown (6) as others were difficult to place due to the distributions of the various taxa and the number of species involved ($\approx$ 100). Male-haploidy has evolved independently at least twelve times (Borgia, 1980), with this expected to increase with future research.

tissue haploid (in addition to germline cells). It is subsequently with the embryonic PGE and arrhenotokous systems that I derive the definition for haplo-diploidy.

How I define haplo-diploidy/male-haploidy

In lieu of multiple, different possible definitions, I attempted to use a simple definition for male-haploidy that includes multiple different mechanisms of male-haploidy (arrhenotoky, embryonic PGE). The definition that I chose follows the mechanisms shown in Figure 2. The definition

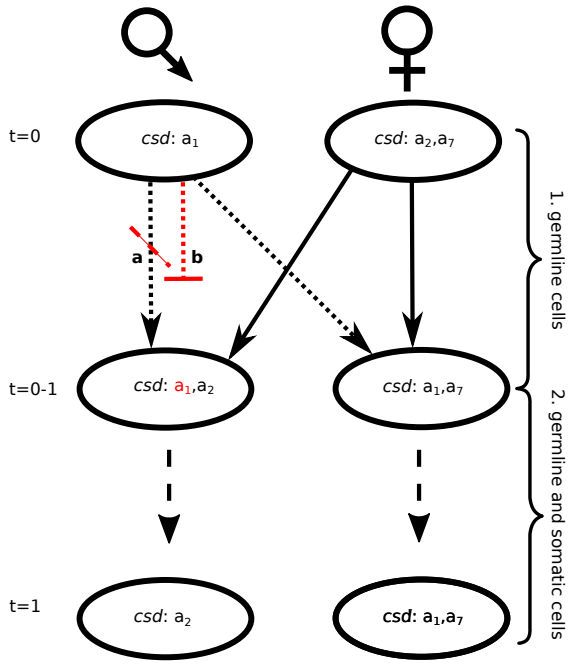

Figure 2: **A reproductive schematic functionally equivalent to either arrhenotoky (b), or PGE (a)**. In arrhenotoky eggs from the female remain unfertilized (**b**), leading to the development of haploid males. With PGE the paternal genome is transmitted (**a**, $t = 0 - 1$) but is lost early in zygote development and does not appear in either germline or somatic cells ($t = 1$). If the mother ($t = 0$,*csd*: a_2a_7) mates with her male offspring ($t = 1$,*csd*: a_2), 50% of the fertilized offspring would develop as diploid males. In contrast females develop normally from fertilized eggs under PGE and arrhenotoky.

incorporates selection as occurring on both haploid and diploid phases: this translates to male-haploidy being expressed in both somatic and germ-cell lines. In terms of reproduction, the method is as follows. For male-haploid populations, *a*) female offspring derive from successfully fertilized eggs and retain the full chromosome complement; *b*) male offspring either: *i*) develop from unfertilized eggs, or *ii*) develop from fertilized eggs but lose the male chromosomes early in development; *c*) male and female members undergo selection; *d*) mutation is included before reproduction. The choice over whether an egg develops either as a female, or male is decided randomly via environmental sex determination, a commonly utilized approach (Goldstein, 1994; Smith, 2000; Engelstaedter & Hurst, 2006).

Part II: Theories on the origin of haplo-diploidy

Background

Historically it has been suggested that PGE and arrhenotoky respectively represent a primitive and a derived state according to Hartl & Brown (1970). More specifically the Schrader

and Hughes-Schrader hypothesis described a four step process through which arrhenotoky could evolve. Under this system PGE could be represented by the same steps as arrhenotoky, with the exclusion of unfertilized eggs being able to develop into males. Although the work by Hartl & Brown (1970) reported that these studies did not provide a suitable theory for the evolution of arrhenotoky. This is despite the later conclusion of Bull (1985) that these theories posed no problems in regards to their viability according to selection on sex-ratios.

A more consistent problem in the evolution of male-haploidy is the viability of haploid individuals. The viability of haploid individuals being a product of at least four factors (Bull, 1985): *i*) The cost of unfertilized eggs developing as males, *ii*) the potential for gene-dosage issues, *iii*) the continued operation of spermatogenesis in haploid males and *iv*) the hemizygotic expression of deleterious alleles. Responses to two of these issues have been found in the natural history of male-haploidy, in particular the ecological niches occupied by their ancestral groups: One of which is that a lot of these (male-haploid) lineages derive from semi-social consumers of "woody tissue" (nutrient deficient) (Normark, 2003). Conceptually this niche, semi-social with nutrient poor diets, was expected to favour inbreeding and the evolution of endosymbionts (both of which factor in with the evolution of male-haploidy (Kuijper & Pen, 2010)). Endosymbionts, if they are maternally inherited and if males and females are in direct competition, will benefit by killing males (Normark, 2004; Werren & Beukeboom, 1998), as males represent an evolutionary "dead end". This endosymbiont killing of males leads to a conflict between the cytoplasmic 'genome' and the nuclear. This may lead to PGE if the nuclear genome evolves to increase the haploid male viability (Normark, 2004; Úbeda & Normark, 2006). Inbreeding can favour haploidy through two mechanisms: *i*) mutations are more likely to be homozygous in highly inbreeding populations, than outbred. Hence, the hemizygotic expression of deleterious alleles is more, or less equivalent to their homozygotic expression in their ancestral diploid population. *ii*) Inbred populations typically have stronger population structures than outbred populations. Hence, inbred populations should, typically, feature less competition between their haploid and diploid males.

Following Hartl & Brown (1970) the current state of knowledge suggests alternative theories for the evolution of male-haploidy (Engelstaedter & Hurst, 2006; Asplen et al., 2009). From a theoretical viewpoint the conditions conducive to these earlier theories have been shown to be limited (Engelstaedter & Hurst, 2006). Furthermore, in some cases where this hypothesis has previously been cited for (Bull, 1985), have more recently been shown to lack one or more of the required conditions (Asplen et al., 2009). As a consequence of this and some investigations using cytoplasmically inherited agents governing ploidy (Otto & Marks, 1996), I used more generalised hypotheses for male-haploidy.

Investigated hypotheses

Chronologically the maternal transmission hypothesis (MTH) and the deleterious mutation hypothesis (DMH, (Hartl & Brown, 1970)) are relatively old hypotheses for male-haploidy. (The

genomic imprinting hypothesis coming considerably later (Beukeboom, 1995)). One of the first formal definitions for MTH can be found in (Bull, 1979; Borgia, 1980); the core premise of which is that females with haploid sons are better represented in future generations than their diploid bearing compatriots. While diploid bearing sons may transmit either paternally or maternally derived genes, haploid sons will transmit only their maternally derived genes. Hence, females bearing haploid sons will be better represented (genetically, through their sons) in subsequent generations. The DMH in contrast relies on indirect selection and linkage disequilibria (Goldstein, 1994); With male-haploidy deleterious mutations will be observed at the rate q (q symbolizes the frequency of deleterious mutations in traditional population genetics), depending on the frequency of males. In diploids deleterious mutations will be observed as often as in females, $q^2 \times 100$ percent of the time. Subsequently, associations between male-haploidy and deleterious mutations will be lower after selection; In the DMH this association of male-haploidy with fitter alleles (p, the wild-type allele) leads to indirect positive selection for male-haploidy.

Analyses of the MTH and DMH form the subject of my first chapter. This decision was based on the observation that both the MTH and DMH will be relevant for any form of male-haploidy (as we describe it). Other hypotheses, such as the male-killing hypothesis, are related to specific biological mechanisms. One exception to this is the theory from Bull (Bull, 1985), regarding sex-ratio selection, which they found to be compatible with male-haploidy. However, with the sex-ratio selection theory being compatible with the evolution of male-haploidy (Bull, 1985), we focused here on the MTH and DMH (also to drive down the number of parameters). The aim of this study was to measure the differences between these two hypotheses, and the impact of mathematical/probabilistic models.

Part III: Impact of haplo-diploidy on host-parasite interactions

Switching between different sex determination methods, diplonty (2N-2N) to haplo-diploidy (2N-N), would be expected to have a large impact on fitness at the species and population level. However, as previously mentioned, this switch seems to happen rather frequently, belying much of a cost in fitness. Might it be possible that most mutations have co-dominant effects (Charlesworth & Charlesworth, 1998), and so infer a weaker cost when homo-, or hemi-zygously expressed? Or, are haplo-diploid systems more robust to the concept of "genetic load" (Hedrick et al., 1997); exhibiting lower inbreeding-depression under experimental conditions (Henter, 2003). Or, is it similar to the idea of genetic purging, more embryos are lost as a result of genetic load, but those surviving development are "purged" somewhat (Hedrick, 1994). Adopting the latter would indicate that a switch between diplonty to haplo-diploidy would result in a new selection-mutation equilibrium for genetic diversity. It could also mean changes in the efficacy of selection arising from species interacting with their biotic environment.

In regards to host-parasite interactions, the effect on fitness of switching is dependent on the degree of variation in infection explained by genetics. If the degree of resistance to be being parasitized is strongly determined by genetics then the effect of switching systems will be

correspondingly strong. Generally it is thought that a significant degree of variation in resistance is determined by genetics (Møller, 2006). The effect on host-parasite interactions is, however, far from obvious; multiple effects from changing the sex-system will have an effect on these co-evolutionary relationships (Zayed, 2004). One of the most ignored impacts in host-parasite co-evolutionary literature is that on dealing with the degree of susceptibility in host populations (Lively, 2010); a property that will change as haplo-diploidy arises in diplontic populations.

Changes in the dynamics of host-parasite co-evolution can have far reaching effects according to the "Red Queen Hypothesis" ("it takes all the running you can do, to keep in the same place" (Carroll, 1960)), this extends to sex and recombination. The Red Queen Hypothesis (RQH) derives from the paper by Hamilton (1964), refinements of which have been made by Barton (1995); Engelstädter & Bonhoeffer (2009) and many others (Otto & Nuismer, 2004; Salathe et al., 2008; Moran et al., 2011; Vergara et al., 2014). In these works the conditions favouring sex have been identified to require additional factors; the RQH alone cannot explain the ubiquity of sex (Otto & Nuismer, 2004). Howard & Lively (1994) showed that host-parasite co-evolution in conjunction with mutation accumulation can better accommodate the ubiquity of sex. Hodgeson & Otto (2012) demonstrated a similar effect from including directional selection.

Interest in the RQH and the evolution of diplonty and haplonty has been expressed by Otto & Marks (1996). Haplo-diploidy (here referred to as male-haploidy) as an intermediate form, has however not been previously studied; An interest of which is present for the role of sex and recombination in the large order of the Hymenoptera, composed of over $150,000$ species. First an understanding of host-parasite co-evolution from a theoretical population genetic approach is used. After this we use an actual model for parasite transmission within and between host generations. This second aspect being frequently ignored in previous studies (Salathe et al., 2007; Agrawal, 2009), and suggested in some recent reviews (Lively, 2010).

Part II

Investigations into the emergence of male-haploidy

Chapter 1

Conditions for the invasion of male-haploidy in diploid populations

Abstract

Male-haploidy has independently evolved several times in different phylogenetic groups and has led to various extant lineages in the Insects, Arachnida and Rotifera. While the stability of male-haploidy as an evolutionary strategy is not well understood, various theories address the invasion of male-haploidy in diploid populations. Here two of these theories: *i*) the Maternal Transmission Hypothesis (MTH) and *ii*) the Deleterious Mutation Hypothesis (DMH), are re-investigated with an agent based model to understand the role of genetic drift as a mechanism facilitating the spread of male-haploidy. These two hypotheses are analysed separately and comparatively and the results suggest dominance of the MTH. In addition, comparison of the stochastic results to deterministic results using the same model structure show how genetic drift can enhance the parameter space where male-haploidy can be expected to invade.

1.1 Introduction

The evolution of organisms with diploid genomes is one of the major transitions in evolution (Szathmary & Maynard, 1995). Diploidy is considered to provide a variety of fitness benefits in comparison to organisms with haploid genomes. In spite of these fitness benefits many species evolved where one sex is haploid and the other diploid. It seems clear that the evolution of diploid organisms predated the evolution of these haplo-diploid species, which then must have been derived from diploid ancestors. These haplo-diploid species are present in a wide range of

invertebrate lineages including the Acari, Hymenoptera, Coleoptera and Diptera (Normark, 2003, 2006). The main mechanisms to achieve male-haploidy are either arrhenotoky (parthenogenetic production of males), or paternal genome elimination (PGE). The actual cytological process can be manifold and any particular instance of the evolution of male-haploidy could be ascribed as the result of several of many hypotheses (Haig, 1993; Ross et al., 2010, 2012a); with manifold impacts on species ecology (de la Filia et al., 2015). Two hypotheses, which are based on basic genetic principles and can be broadly applied over the majority of male-haploid cases, are: *i*) The maternal transmission hypothesis (MTH), which is based on the female biased nature of inheritance inherent to male-haploid systems. And *ii*) the deleterious mutation hypothesis (DMH), which involves indirect selection for the male-haploid allele. The MTH can be easily described through the maternal and paternal inheritance of alleles: The alleles in a heterozygous mother are better represented in two generations time through haploid rather than through diploid sons, since the latter also pass on paternal genetic material. The DMH is also described through alleles: It relies on purging alleles bearing deleterious recessive mutations by "enhanced" selection in haploids. These two hypotheses are applicable to those instances of PGE occurring in the Coleoptera, Diptera and Hemiptera; and to arrhenotoky occurring in the Coleoptera, Hymenoptera, Thrips and Hemiptera. The two hypotheses (MTH and DMH) will here be tested in a model of male-haploidy drawn from natural systems. With a fixed sex ratio of 50:50 to reduce model complexity and ease analysis.

We take our model of male-haploidy from examples in the Diptera, Coleoptera and Hemiptera. PGE species from Coleoptera and Hemiptera require fertilization from the male parent for embryo development, though the paternal genome is subsequently eliminated in male lines. It is from these cases, where the somatic and germ-line cells are exposed to haploidy, which we take our model for male-haploidy. In contrast, the Sciaridae model of paternal genome elimination, where the paternal genome is lost in the germ-line (Haig, 1993), is not in keeping with the model and beyond the scope of this study. Arrhenotoky from the Coleoptera, Thrips, Hymenoptera and Hemiptera with the maternal determination (Sasaki & Obara, 2002) and parthenogenetic production of males is less in keeping (Normark, 2003), due to the sex-determination mechanism: In the Hymenoptera sex-determination depends either on a single locus (the *csd* locus, complementary sex determining) (Stahlhut & Cowan, 2004), on multiple loci (Crozier, 1971; Miyakawa & A.S., 2015; de Boer et al., 2015) with multiple alleles, or genomic imprinting (Beukeboom & van de Zande, 2010); Fertilized heterozygous individuals develop as females, and hemi-, or homozygous individuals as males. In some cases fertilization of the eggs has been demonstrated to be under maternal control (Gerber & Klosterm, 1970; Sasaki & Obara, 2002). In our model paternal genomes are either: *i*) not expressed in males that inherit the male-haploidy allele from their mother, or *ii*) not transmitted to males that inherit the male-haploidy allele from their mother. Noting that sex determination, in our model, did not fall under maternal control. In either case the paternal-portion of the male genome will be neither phenotypically expressed, nor heritable in sons, independent of *i*, or *ii*.

Using the model for male-haploidy we will explore the ability of this system to invade a traditional diploid population; furthermore we can explore the consequences for a population that is under invasion by male-haploidy. Invasion by male-haploidy is expected to have a large impact on the evolution of any genome, foremost by the expression of all recessive mutations in the hemizygotic individuals. The immediate result is a much higher overall population fitness cost, as deleterious alleles are expressed at the rate of q rather than at q^2, as for diploids. The knock-on effect of this is the loss of genetic load, as selection is more effective on haploid genomes. Further impacts are expected through the absence of recombination (in the haploid gender), and the reduction in effective population size; both being equivalent to the effect of the X-, or Z-chromosomes. So how does the fitness benefit of removing the genetic load from a population compare to the costs of evolving male-haploidy in the first place, and the reduction in effective population size?

1.2 Methods

The general model

The simulations are based on the invasion of male-haploid mutants into a diploid population. Male-haploidy is considered neutral and selection operates on a viability locus not affecting and independent of ploidy. Hence the simulations begin with diploid populations, containing both wild-type and deleterious mutations (with respectively corresponding fitnesses of: $w_p = 1$ and $w_q = 1 - s$; and heterozygotes $w_{pq} = 1 - hs$, where $h = 0$) at the viability locus. Regarding the viability locus, $f(p)$ was used to indicate the frequency of the wild type allele upon initialisation, while $f(q)$ corresponded to the deleterious mutant. The wild-type allele was assumed to be free of deleterious mutations and was given the fitness 1. The fitness cost of the deleterious mutants was provided by the selection coefficient (s) and the heterozygote effect (h). Mutation between these alleles occurred at rate ν from the wild- to mutant-types, and at μ for the reverse rate.

In addition we included a single locus controlling male ploidy to which we refer as the "modifier" locus. Two alleles at this locus control male ploidy: D for diploid males and d for haploid males, occurring at population wide frequencies of $f(D)$ and $f(d)$, respectively. (This locus [**D**] being unrelated to the sex determination process). We also tested conditions where this modifier locus incurred a direct cost for male-haploidy (d is associated with a fitness cost, c). We introduce this additional cost due to the manifold consequences of male-haploidy. For arrhenotoky to evolve: *i*) unfertilized eggs need to be able to develop into viable adults, *ii*) haploid males need to be able to produce sperm, *iii*) gene-dosage issues, and *iv*) the consequences of genetic load. For PGE items *ii*, *iii* and *iv* are present, the viability locus (**P**) only accounts for item *iv*. Hence this cost (c) is used to account for items *i*, *ii* and *iii*, and was independent from the viability locus. In addition, the modifier locus was considered to have no direct phenotypic effect in females, the ploidy of male offspring being an indirect effect. Variation in the modifier locus' frequency and distribution was subsequently dependent on its inheritance and phenotypic effects in males.

In the model these parameters were organised within vectors, or matrices; as appropriate: The genotype frequencies ($f(D^2p^2), ...$) and selection on the viability locus ($..., 1-hs, 1-s$) were stored in vectors. These frequencies are henceforth recognised by the symbols $Y1, Y2, ...Y5$ for the male and $X1, ...X9$ for the female genotypes (See the supplementary material Frequency vectors). While the mutation rates were stored in matrix form. Calculation of the next generation frequencies involved the use of inheritance matrices ($(\mathbf{X1}, ..., \mathbf{X9}) = \mathbf{H^f}$), organised within the vectors ($\mathbf{H^f}$ and $\mathbf{H^m}$). The inheritance matrices for the individual genotypes can be seen in the supplementary material.

The initial population genotype frequencies were $\approx f(X1), f(X2), f(X3)$; $f()$ proceeded by parentheses indicate frequencies of the respective elements. Where D is set to ≈ 1, such that $f(1-D^2)$ is effectively 0 upon initialization in the stochastic simulations. $\mathbf{Q^f}$ is used for the genotype frequencies that correspond to the vector containing the female genotypes. $\mathbf{Q^m}$ corresponds to the respective vector of the male genotypes. The frequencies of the male vector then becomes:

$$\mathbf{Q^m} = \begin{pmatrix} f(Y1), f(Y2), f(Y3), f(Y4), f(Y5) \end{pmatrix}$$

In the selection vectors overdominance is included ($h = -1$) and allowed to vary to include co-dominance and complete dominance ($h = 1$). The vector of selection coefficients for males also included the cost of being haploid (c), this is further illustrated in the selection vectors section of the supplementary material. This male-viability cost (c) can be translated to the requirement of the health of haploids to be at least half that of diploids to invade in previous analyses (Smith, 2000). The vector of the various selection coefficients for males is given as follows:

$$\mathbf{S^m} = \begin{pmatrix} 1, (1-hs), (1-s), (1-c), (1-c)(1-s) \end{pmatrix}$$

Calculation of the genotype frequencies after selection used the divergence of the individual genotypic fitnesses from the population average. The divergence from the population average fitness is provided by correcting the elements of the selection vector ($\mathbf{S^f}$) with the sum of the product of genotype frequencies by their fitness effects ($\sum_i \mathbf{Q}_i^f \mathbf{S}_i^{f\mathrm{T}}$).

$$\begin{aligned} \mathbf{Q}^{f'} &= \mathbf{Q}_i^f \frac{\mathbf{S}_i^f}{\sum_i \mathbf{Q}_i^f \mathbf{S}_i^{f\mathrm{T}}} \\ \mathbf{Q}^{m'} &= \mathbf{Q}_i^m \frac{\mathbf{S}_i^m}{\sum_i \mathbf{Q}_i^m \mathbf{S}_i^{m\mathrm{T}}} \end{aligned} \tag{1.1}$$

Using mutation matrices involves recycling the use of the frequency vectors ($\mathbf{Q}$) over the rows of the mutation matrix ($\mathbf{M^f}$ for females). For each row, the sum of the product of the elements from the frequency vector and row elements of the mutation matrix are taken to provide the new frequencies after mutation. Where each row (j) of the mutation matrix corresponds to the respective element (j) of the frequency vector ($\mathbf{Q}_\mathrm{j}$).

$$\begin{aligned} \mathbf{Q}^{f''} &= [\mathbf{M}^{f'}\mathbf{Q}^{f'\mathrm{T}}] \\ \mathbf{Q}^{m''} &= [\mathbf{M}^{m'}\mathbf{Q}^{m'\mathrm{T}}] \end{aligned} \tag{1.2}$$

The next generation frequencies were calculated by constructing inheritance matrices (**X1** for the female wild type inheritance matrix). These inheritance matrices were structured with the column elements corresponding to the female genotypes, and the row elements to the male genotypes. Of the elements within the matrices each corresponds to the proportion of matings expected to result in the i, j^{th} genotype. The inheritance matrices under assortative mating take two forms **H** and **A**; non- and assortative respectively. These are then combined according to the parameter A. The next generation frequencies then become the combined sum of the product of all elements within the matrices (Equations 3 & 4). Under the maternal inheritance rules used here, the inheritance matrices for haploid males only feature values greater than zero when females contain the d (mutant male-haploid causing) allele: Haploid males are determined purely from maternally inherited genetic elements.

$$\mathbf{Q}_i^f = (1-A)\dot{[}\mathbf{Q}^f[\mathbf{Q}^{f\mathrm{T}}\mathbf{H}]_i^f]_j + A\dot{[}\mathbf{Q}^f[\mathbf{Q}^{f\mathrm{T}}\mathbf{A}]_i^f]_j \tag{1.3}$$

$$\mathbf{Q}_i^m = (1-A)\dot{[}\mathbf{Q}^m[\mathbf{Q}^{m\mathrm{T}}\mathbf{H}]_i^m]_j + A\dot{[}\mathbf{Q}^m[\mathbf{Q}^{m\mathrm{T}}\mathbf{A}]_i^m]_j \tag{1.4}$$

The stochastic model

The model structure was the same retained for the stochastic models. Adding stochasticity involved introducing a rejection sampling approach for randomizing the effects of selection, mutation and genotypic assortment (Beaumont et al., 2002). In addition to the above, the population size within the stochastic simulations was allowed to vary from 500 thousand to 20 million individuals.

Analysis

The parameters through which we explore this process include the incorporation of overdominance ($-1 \leq h \leq 1$) through to dominance of the allele under selection. The expression of selection, from absence/weak ($0 \leq s < 1$) to lethal. The de novo costs of male-haploidy ($0 \leq c < 1$), and assortative mating ($0 \leq A < 1$). In addition variations in the mutation rates are also included (10E^{-06} to 10E^{-03}) for both forward and back mutations. Illustration of these results were performed in both R-cran v2.13.1 (R Core Team, 2013) and gnuplot (Williams & Kelly, 2003).

1.3 Results

The male-haploid allele invades the diploid population under a wide range of scenarios, including mildly deleterious and recessive mutations assumed to be representative of conditions in natural

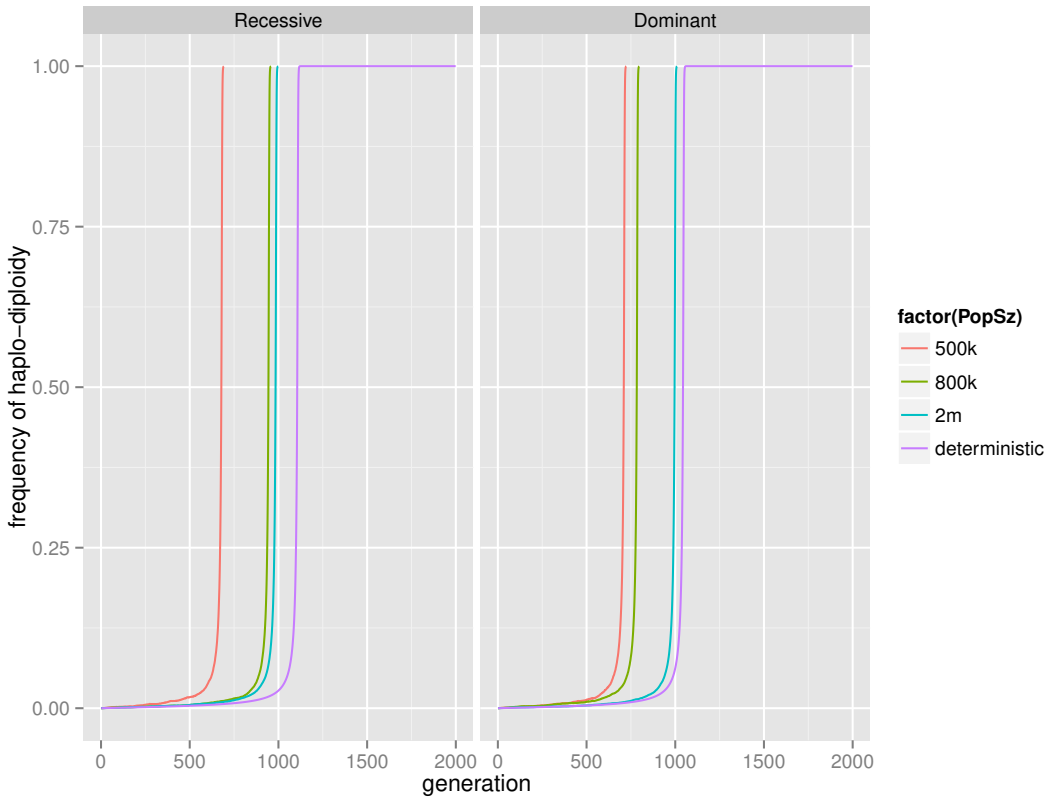

Figure 1.1: **The number of generations required until fixation of the male-haploid modifier allele**. Conditions vary from recessive to dominant mutational effects in deterministic and stochastic simulatiom models. Under stochastic models the absolute population size varies from 500k to 2m, showing the variance resulting from differing degrees of stochasticity. The smaller population sizes correlate with the earlier invasion of the modifier allele.

populations. The forward and reverse mutation rates were given as $\mu = \nu = 10E^{-03}$ for the viability locus in Figure 1. Mutation rates for the modifier locus (determining ploidy) were set at $\mu_d = \nu_d = 10E^{-06}$, the selection coefficient was kept at $s = 0.05$ in both subplots. With recessive mutations on the viability locus, selection on the modifier locus is negative. Nevertheless, in spite of selection against the mutant allele being stronger, the mutant allele was observed to spread in the simulation results (Figure 1). Fixation of the d allele at the modifier locus occurred around the 700^{th} generation mark under recessive mutations and a population size of one million.

In the case of a completely dominant mutant allele q, the d allele at the modifier locus comes under positive selection. Subsequently the number of generations required before the d allele (for male-haploidy) becomes fixed, is markedly lower than with a recessive mutation at the viability locus (Figure 1). Changes in the degree of stochasticity (increasing with smaller populations, $N* < N$) correlate with the "time until fixation" of the d allele. The greater degree

of stochasticity in the model is associated with a more rapid process of fixation of male-haploidy (Figure 1).

Figure 1 remains representative of the situations with regards to positive values of h. If the mutant allele q follows the expression pattern of overdominant alleles (heterozygous individuals have a greater fitness, $h < 0$) selection ($0.1 < s < 1$) acts more strongly in favour of the diploid allele (D, Figure 2). Subsequently the male-haploid allele (d) either remains at low frequency, or is effectively lost.

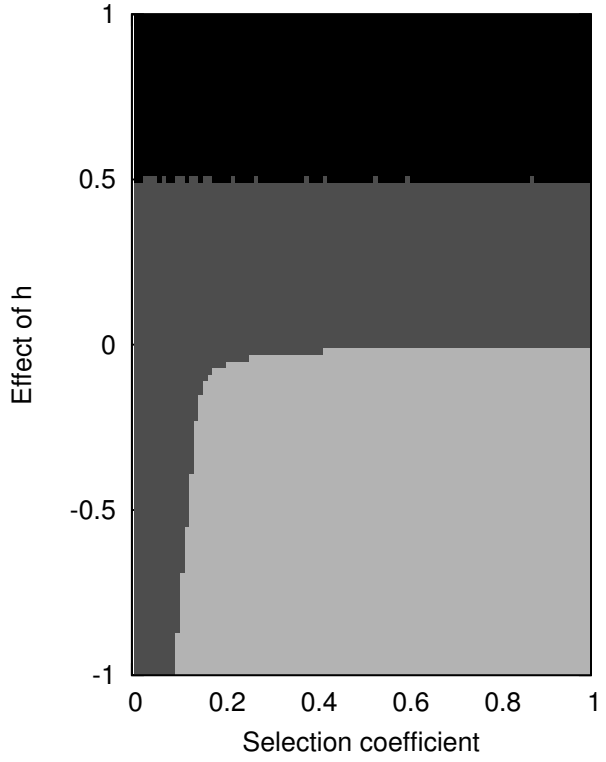

Figure 1.2: **The range of parameter values where the male-haploid allele (d) reaches fixation (the black and darker gray shaded regions, $0.0 < h < 1$ and when $s < 0.1$).** Upon fixation of the male-haploidy term the population fitness increases by 0.0095 - 0.09125 (data in supplementary material), unless the mutation is strongly dominant (the black shaded region, $0.5 < h < 1$). When selection remains weak (< 0.1) the male-haploid allele reaches fixation regardless of the over-dominant/recessive/dominant nature of selection. Under this region (the darker gray shaded area), the male-haploid allele becomes fixed despite being weakly selected against. With stronger selection and over-dominance the male-haploid allele fails to reach fixation (the lighter gray region).

Upon introduction of *de novo* costs (parameter c), reductions in the fitness of male-haploidy by $\frac{1}{20}$ (relative to diploids) is enough to severely curtail the region of parameter space in which male-haploidy can invade. These conditions change under assortative mating on the viability locus; with assortative mating invasion occurs over a greater range of male viability costs ($0 \leq c < 0.5$,

Figure 3).

Under higher mutation rates at the viability locus (both forward and backward mutation rates from p to q) the region of parameter space where male-haploidy spreads is enhanced. Diploidy is maintained under either extreme overdominance, strongly deleterious mutations, or moderate to high costs in male-haploid viability. Male-haploidy may continue to be favoured even under these adverse conditions if assortative mating is introduced to the model. Further, introducing stochasticity increases the parameter range where male-haploidy can invade, excluding the male-viability costs (Figure 4). Alternatively, incorporating stochasticity reduces the stability of diploidy against invasion by male-haploidy, over a greater parameter range.

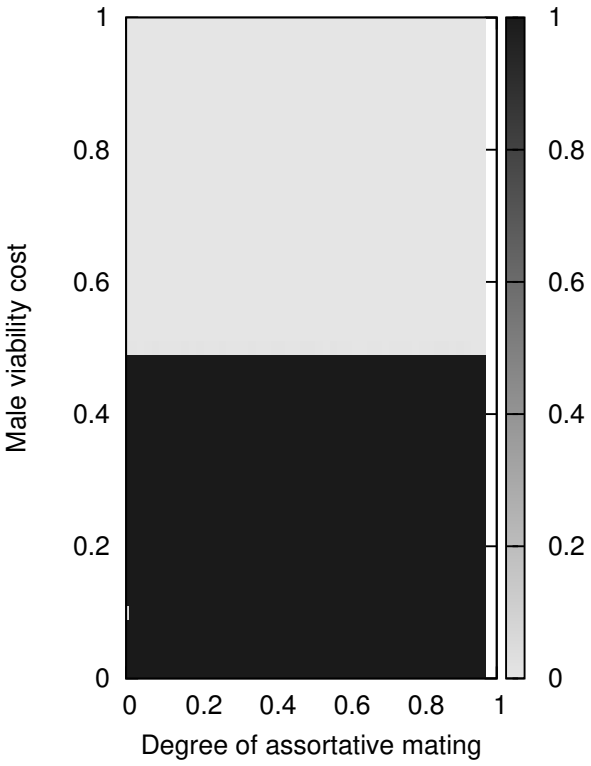

Figure 1.3: **The relationship between the degree of assortative mating and the viability cost of haploid males**. The darker regions of the plot (values ≈ 1 on the colour scale) indicate the region of parameter space where male-haploidy invaded. In the absence of assortative mating male-haploidy only invades under a very reduced range of male-haploid viability costs (< 0.04). Under assortative mating a much larger range of male-haploid viability costs support invasion, with this showing little sensitivity to the degree of assortative mating.

1.4 Discussion

The simulation results from the deterministic model diverge from the previous results of Goldstein (1994) and Smith (2000): the parameter range conducive to the spread of the modifier allele (the allele that modifies the ploidy state of males, from diploid to haploid) is wider (Figure 2). Divergence of the results from those of Goldstein (1994) arise from the inclusion of the MTH, which he explicitly excluded. The discrepancies between our results and those of Smith (2000) may result from: *i*) Smith (2000) considering the cost of haploid males as a relative viability compared to diploids, whereas we consider it as a fixed cost *ii*) Smith (2000) measuring invasion of the modifier allele (d here, P in Smith (2000)) as penetrating diploid populations, whereas here successful invasion is reported when the allele reached fixation. Hence while we record that a fixed male viability cost of 0.04 on haploidy prevents fixation (the modifier allele increases to between 0.3–0.6), this may still be equivalent to fixation under Smith (2000), depending on the rate of change in the modifier. A further cause for divergence is the consideration of the male-haploidy locus being dominant (when maternally inherited). This has a large impact as under this system males cannot be heterozygous at the male-haploidy locus.

The results from the stochastic simulations diverge further from the results of Smith (2000) and Goldstein (1994). Stochasticity was introduced into the model via a probabilistic approach to calculating the effects of selection, mutation and reproduction, barring recombination. Hence, the introduction of stochasticity can be partially interpreted through the prism of genetic drift. The most striking impact is the reduced sensitivity of the invasion of male-haploidy to the s and h parameters (Figure 4 compared to Figure 2). Without a direct cost acting upon the modifier allele, male-haploidy spreads to fixation under all combinations of s and h. In contrast, the results regarding the direct cost (c) in the deterministic (not shown) compared to probabilistic model (Figure 4) show little to no variation. The increased stability of invasion by male-haploidy to indirect, but not direct selection indicates the role of stochasticity and genetic drift on allele frequencies. Stochastic processes might either cause: *i*) a less effective selection process, or *ii*) more realistic genotype frequency distributions. A key problem with the selection argument is that the population sizes were far too high to let stochastic sampling render selection at the viability locus "neutral" (when $N_e s \approx 1$). The indirect selection coefficients in the simulations were also consistently several orders of magnitude greater than the threshold for neutrality. In contrast the genotype frequency distributions may be more important for several reasons: Firstly, the genotype frequencies are partially-independent model variables (dependent on previous values, selection and drift) in contrast to selection as a fixed model parameter. Partially-independent model variables are far more liable to effects from stochastic events, unlike fixed model parameters. Secondly, the MTH is less dependent upon selection than the DMH, and has been shown to be dominant to the DMH (Smith, 2000).

The direct cost, as previously mentioned (c), derives from the expectation that male-haploid genes are more susceptible to selection on genetic load (as for X-linked genes). This expectation has support from experimental studies showing that X-linked genes are under lower genetic load

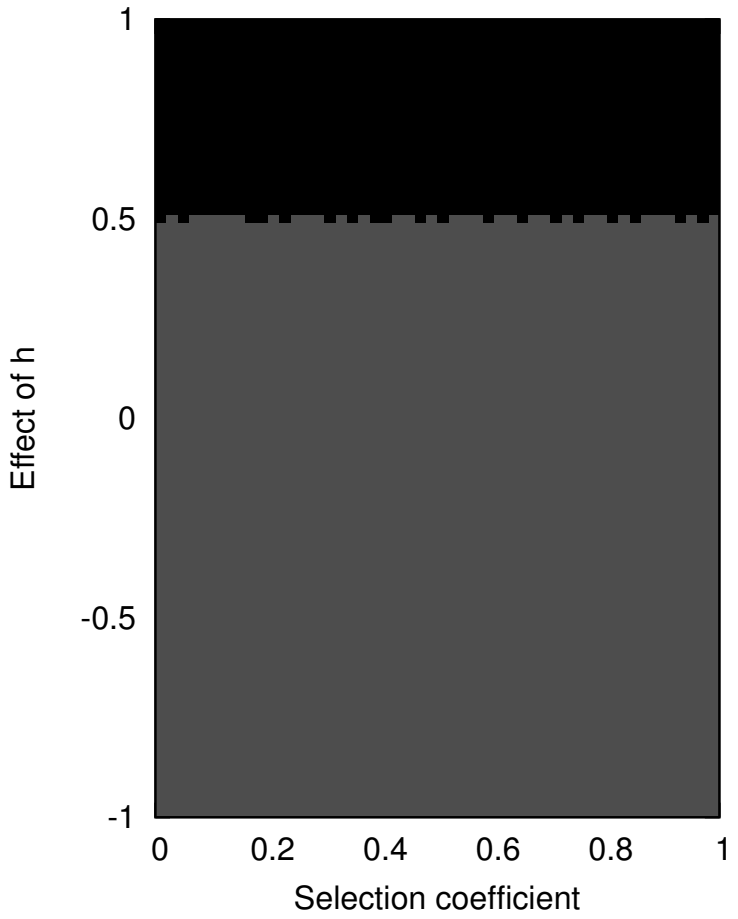

Figure 1.4: **The results from running stochastic simulations over the same parameter space as in the deterministic model**. The region of parameter space where male-haploidy invades is extended over those conditions when it remains under negative selection pressure (the dark gray region, compared to figure 2).

than diploid autosomal genes (Hedrick et al., 1997). Whether a sex-linked X-chromosomal model can be easily expanded to a genome-wide model is not clear, hence reasonable costs for male-haploidy remain unavailable. Consequently the viability cost for male-haploidy (c) was added. Upon introduction of this parameter the invasion of male-haploidy became dependent on the presence of assortative mating (Figure 3). While the degree of assortative mating had a large impact on the duration of the invasion process, no impact was observed for the probability of invasion. Hence, no interaction between the degree of assortative mating and the viability cost in Figure 3. (Male-haploidy did not invade if the viability cost is greater than 0.5, as the advantage conferred by the MTH is removed.) This result is the same as those previously found in similar studies (Goldstein, 1994; Otto & Marks, 1996; Smith, 2000).

Maternal transmission versus deleterious mutation hypothesis

It is possible to elucidate the relative contributions of the deleterious mutation (DMH) and maternal transmission (MTH) hypotheses by correcting the offspring frequencies associated with the male-haploid allele (d). Let us consider a female, heterozygous at the male-haploid determining locus. Male offspring inheriting the male-haploid allele will always contribute the maternal genetic material to their offspring. In contrast the diploid male offspring have a 50% chance of transmitting the maternal material to each female offspring. We correct for this inheritance bias by multiplying the diploid male frequencies by $\frac{4}{3}$ and the male-haploid by $\frac{2}{3}$ (see Figure 5). With the correction of the male-haploid haplotype frequencies the male-haploid allele failed to invade ($g < 0.01$). This suggests that the MTH is the dominant process driving the results here, as in previous studies (Smith, 2000). However, in the absence of assortative mating the current simulations are highly sensitive to direct costs on the male-haploid allele: If the male-haploid allele incurs a direct disadvantage of $\frac{1}{20}$ compared to diploid individuals it fails to supplant diploidy, although invasions still occur. These observations are independent of whether the cost is borne by either the females, or males bearing the male-haploid allele.

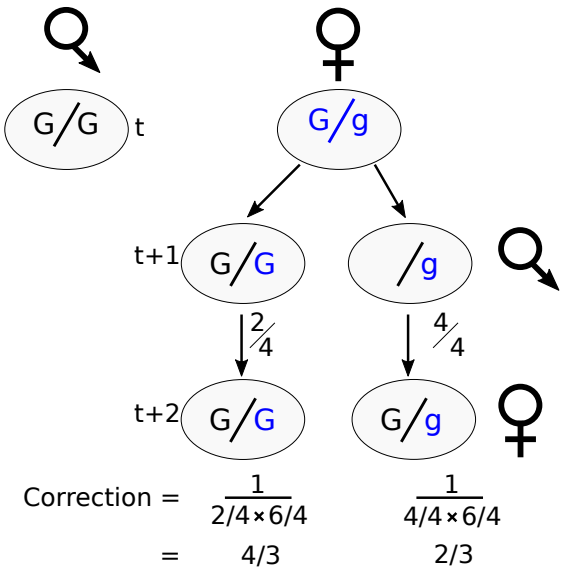

Figure 1.5: **The illustration of the Maternal Transmission Hypothesis (MTH) over 3 generations demonstrates the inherent advantage of male-haploidy over traditional diploidy**. Female offspring from a mothers son confer twice as much genetic material inherited from their mother when they are haploid. This advantage can be removed by adjusting the male frequencies at $t+1$, effectively halving the number of haploid males that a heterozygous mother produces.

The different contributions from the DMH and MTH have an impact on the role of effective population sizes on the estimated time until fixation. This difference arises from the different rates at which both processes are expected to operate (Smith, 2000). The DMH will occur at the rate at which deleterious mutations are expected to arise within the population. In contrast the MTH occurs at a rate dependent upon the sex ratios: A female heterozygous at the male-haploidy locus, given the adequate supply of sperm, produces a $\frac{50}{50}$ ratio of haploid/diploid males, independent of the male genotype. Hence the frequency of male-haploidy in male offspring is potentially double that of their parents, leading to a higher rate of invasion compared to the

DMH. The results observed here generally concur with those from Smith (2000), but both results suffer from the same limitations (a single locus viability model and arbitrarily imposed costs). This convergence in the results between our model and those of Smith (2000) come despite the use of different reproductive models. While the model of Smith (2000) resembles arrhenotoky, ours more closely resembles somatic PGE. This results in two main differences: *i*) "mutant" alleles at the modifier locus being recessive (Smith, 2000) versus co-dominant, *ii*) males being heterozygous at the modifier locus (Smith, 2000) versus their exclusion. The first will result in a stronger effect from the MTH, due to the distribution of female matings that will lead to the MTH (from d^2 to d). In the absence of male heterozygotes (at the modifier locus) an additional cost may arise through the availability of viable mates for producing male offspring. Without male heterozygotes the PGE model incurs a cost against the modifier allele d. This cost arises from fewer avenues of inheritance through males: Arrhenotokous males, either heterozygous or homozygous for d, can transmit the modifier allele to male offspring. PGE males, in contrast, have no such opportunity for the transmission of the modifier allele d to male offspring. This "viable matings" cost of PGE is low, frequency dependent, and like the MTH dependent upon the sex ratio. Assuming discrete generations this cost should also operate independently of the MTH, because no such cost occurs for their daughters (who are solely responsible for the MTH). Under heterozygote advantage different evolutionary strategies should be expected for arrhenotoky and PGE systems. Given the independence of the MTH from the viability locus (with free recombination, as here), the expected equilibrium frequency for the mutant allele d will be higher for PGE systems. Under dominant to recessive conditions the evolutionary stable strategies are expected to be equivalent. With drift driven changes in frequencies in the stochastic model (as mentioned earlier in the discussion), independence of the modifier from the viability locus, stronger MTH; Male-haploidy may be more likely to invade.

While the methods presented involve a single locus (as in Smith (2000)), in nature the DMH will depend on genome and chromosome-wide rates. Per genome mutation rates are expected to reach around unity (Eyre-Walker & Keightley, 1999), or greater (Nachman & Crowell, 2000). Subsequently the strength of the DMH or MTH may be equivalent under multi-locus models. Adoption of a genome wide model may provide the best solution for identifying the contributions of these two theories to invasion by male-haploidy modifiers.

To better identify the value of the different hypotheses more information is required on several aspects. One of the most important and at the same time least well known is the viability cost of male-haploidy. Understanding the fitness landscape at the genomic level, through mutation, may help to estimate these costs. Plenty of theoretical models are available for this from the population genetics literature (Charlesworth & Charlesworth, 1998), but empirical studies are lacking that actually test the predictions made by theory. Even with more empirical studies understanding why male-haploidy is not more common in the animal kingdom is expected to remain a theoretical pursuit.

Acknowledgements

We would like to thank the referees and editor for their contributions and improvments to this paper. Funding for the research was provided by the Deutsche Furschungsgemeinschaft (DFG), 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) in Halle-Jena-Leipzig, also funded by the DFG (FZT 118). We declare no conflict of interest in the publishing of this manuscript.

Part III

Impacts of male-haploidy on host-parasite theory

Chapter 2

The Red Queen process does not select for high recombination rates in haplodiploid hosts

abstract

One of the main competing theories to describe the evolution of recombination is the Red Queen Hypothesis (RQH). Presently, many theoretical analyses of the RQH typically examine fitness interactions in host-parasite frameworks. Less emphasis has been placed on understanding the impact of host ploidy in these systems. In this study we look to investigate high observed rates of recombination observed in two common haplodiploid species (*Apis mellifera* and *Bombus terrestris*). We compared haplodiploid to diploid host populations under infection with asexual parasites, using a Matching Allele (MAM) model. Results from a simulation analysis showed that the Red Queen does not run in haplodiploid hosts and is therefore, probably not responsible for the high recombination rates observed so far in haplodiploid hosts.

2.1 Introduction

Sex is a costly process and the evolution of sexual recombination is an old puzzle of evolutionary biology (Charlesworth, 1980). Nevertheless, there are purported advantages to sex as posited by the Red Queen Hypothesis (RQH) (Hamilton, 1964). The hypothesis predicts that resistance in host populations infected by multiple parasite strains is maintained through selection for both sex and recombination. The hypothesis requires the presence of negative feedback loops (over generations), whereby common parasite strains select for specific resistance alleles within host populations, which subsequently reduce the prevalence of these strains. This process is similar to negative frequency dependent selection, maintaining genetic variance in both host and parasite populations over generations (Sutton et al., 2011; Takahashi et al., 2011). Sexual recombination may prove beneficial to a host because a genetic association with the most abundant parasite strain would be reduced in subsequent generations. A further potential benefit of sex in the RQH is through the increased amount of genetic variation generated by the reassociation of genetic elements (Barton, 1995; Peters & Lively, 2007).

Currently, predictions arising from the RQH have been tested using either diploid (Schmid-Hempel & Jokela, 2002; Agrawal & Otto, 2006; Agrawal, 2009), or haploid (Peters & Lively, 1999; Otto & Nuismer, 2004; Peters & Lively, 2007; Salathe et al., 2007) host genetic models. In further studies, haploidy was found to be beneficial for parasites (Nuismer & Otto, 2004), whereas higher ploidy levels (diploidy, tetraploidy) were favoured in the host (Oswald & Nuismer, 2007; M'Gonigle & Otto, 2011). The only exception was the evolution of haploid hosts under the MAM (Nuismer & Otto, 2004). The majority of these studies have focused on three interaction models: *i*) the Matching Allele Model (MAM), *ii*) Inverse MAM (IMAM) and *iii*) the Gene for Gene (GfG) model; each of which has its own applicability to certain host-parasite systems. The GfG model was, for example, developed from immunology studies on plants (typically rusts Keen (1990)). Likewise the IMAM was developed on the basis of host recognition and response to parasite infection (an adaptive immune response), while the MAM is based on parasite recognition of the hosts. Since arthropods are not very well known for adaptive immune systems the MAM was adopted as the interaction model over the IMAM and GfG models (Wilfert et al., 2007a) However, since the results of previous studies (Nuismer & Otto, 2004; Oswald & Nuismer, 2007; M'Gonigle & Otto, 2011) suggest the possibility of interdependence between these factors, the results here remain specific to the MAM.

We followed a model system close to that of those arthropods that have been shown to exhibit the highest known recombination rates in Bilateria (social insects, Wilfert et al. (2007a); Bumblebees, Stolle et al. (2011); and Honeybees, Solignac et al. (2004); Meznar et al. (2010)). All of which occur in the order Hymenoptera, an order well known for two features: sociality and haplo-diploidy (Normark, 2003). While most of those species whose genomes have been well studied were either primitively-, or eu-social, sociality is a derived, polyphyletic trait within the hymenoptera whereas haplodiploidy is an ancestral, monophyletic trait common to all Hymenoptera. One of the main theories on the origin of haplodiploidy (allelic penetrance: stronger

phenotypic representation of alleles) may have very strong relevance in regards to the RQH (Smith, 2000). We therefore incorporate the haplodiploid sex determination system in the models presented here. We do not address sociality at this stage though sociality may remain a factor interfering with the evolution of high recombination rates in the social Hymenoptera (differences in recombination rate have been observed between parasitic wasps and eusocial species; Niehuis et al. (2010)).

The model

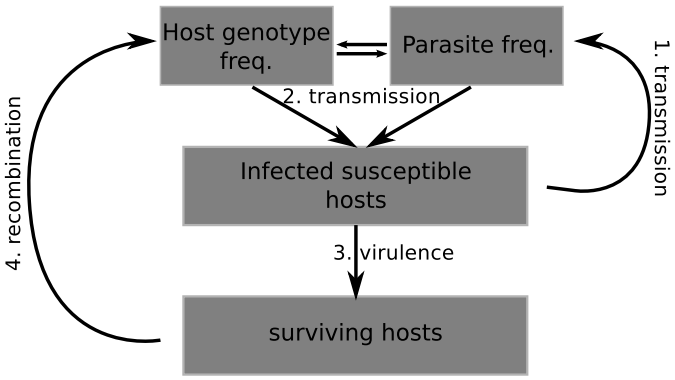

Figure 2.1: **A schematic overview of the model including either a diploid or a haplodiploid host, with a haploid parasite**. The model is based on Hardy-Weinberg assumptions with a deterministic population genetic approach under a two loci, two allele model (noted A/a and B/b). The first transmission step is used to indicate inter-generational transmission, while the second is for intra-generational transmission.

We created both diploid and haplodiploid models, with a haploid parasite under Hardy-Weinberg assumptions. It is based on a simple deterministic population genetic approach with a two loci model (noted A and B Fig 1), with two alleles each (indicated by the upper (A/B) and lower case (a/b). The haploid parasites had also two loci with two alleles and a specific allele combination would allow for an infection of a host with a susceptible genotype. Parasite transmission was based on the frequency of meeting an uninfected susceptible host individual. After infection, selection was calculated per haplotype of the host according to Crow & Kimura (2009) exploring the parameter space for parasite virulence. We excluded dominance effects to simplify population wide estimates of allele selection strengths. Since host individuals can be susceptible to more than one parasite strain we included a transmission process allowing for multiple infections.

To test for the importance of host genome recombination in host-parasite evolution, we included a modifier locus into the model. The modifier locus has two alleles (m and M), initially the population is composed of the first allele (m) with the second allele being introduced. We set the initial genome-wide recombination rate to zero (allele m has no effect on recombination), and increase the size of the effect of the second allele (M) on recombination rates (from 0 to 0.5). The only difference between these two host-parasite systems was the ploidy of the host male genotype.

2.2 Methods

The model used by Engelstädter & Bonhoeffer (2009) was redesigned for use with two host systems (haplodiploid and diploid) using the matching allele model (MAM, Wilfert et al. (2007b). The two interaction loci are referred to as **A** and **B**, while **M** refers to a modifier locus; the alleles were as follows *A/a*,*B/b* and *M/m*. Genetic associations between the loci were calculated in the order of **A** → **B** → **M**. The recombination modifier altered the rate of recombination among all loci, such that recombination rates between **A-B** and **B-M** were equal. This two loci two allele interaction model, two allele modifier locus and asexual parasite population model differed to that of Engelstädter & Bonhoeffer (2009) by the addition of a parasite transmission model.

Parasite transmission occurs during the first two steps of every iteration of the model (counted as host generations). The first step, upon initialisation, was infection of the host population. This was estimated as the frequency of the parasite in the parent population, multiplied by the proportion of susceptible hosts in the offspring (Figure 2.1, step 1). During the inter-generational infection of host populations it was assumed that the parasite did not distinguish between host genotypes, and would attempt to infect all host genotypes randomly. Hence depending on the frequency of resistant host genotypes, a portion of parasite strains failed to infect a host upon each host generation and were removed from the parasite population. Subsequent calculations within the host generation involved probabilities conditional on the proportion of hosts susceptible to the parasite strain (Figure 2.1, step 2). After initial transmission, the within-generation parasite transmission was conducted (Figre 1, step 2), after which selection upon the host was calculated (Figure 2.1, step 3). Finally sex and recombination was modelled within the host population, generating the next generation genotype frequencies to be infecedted by the previously calculated parasite frequencies.

The parasite transmission model required two different processes for within and between generation transmission (Figure 2.1). Within host generations, transmission was calculated as parasite population growth within the susceptible proportion of the host. Between host generations, parasites were transmitted when they infected a susceptible host, with the previous infected proportions used as the probability of contact between host and parasite. Within each host generation the parasite growth rate (T) is calculated as the degree to which uninfected → infected transitions are expected. Therefore, applying the growth rate to the proportion of infected susceptible hosts ($p(P_r)$), gives the new proportion of infected individuals. Transmission to the next generation then requires finding the frequency of susceptible host genotypes in the population (the sum of each host genotype, frequency $Freq(H_i)$, which is multiplied by the proportion infected ($T \times (P_r)$).

$$p(P_r^{t+1}) \leftarrow \mathrm{T} \times \left(p(P_r^t) \times \left(\sum_{i=1}^{i=n} Freq\left(H_i^t\right) \right) \right) \tag{2.1}$$

Fitness of the host genotypes was calculated according to the MAM and the proportion of

the genotype infected by each parasite strain. In calculating selection coefficients it was assumed that having some mis-matching alleles conferred a selective benefit for the host genotype. In diploid genotypes this was incorporated by subtracting $\frac{virulence}{4}$ from *virulence*, while in haploid genotypes $\frac{virulence}{2}$ was subtracted for each mis-matching allele. Incorporating these selection coefficients with multiple infections required matrices identifying all infection combinations: each matrix had n columns and 2^n rows, with the elements over the rows and columns being either one of:

$$\text{proportion surviving infection by parasite strain } r \leftarrow \quad p(P_r) \cdot (1 - s_r) \tag{2.2a}$$

$$\text{proportion of host uninfected by parasite strain } r \leftarrow \quad 1 - p(P_r) \tag{2.2b}$$

c indicating the column (n indicating the number of virulent parasite strains). These matrices and the algorithms used to generate the combinations for these matrices are provided in appendix B. The probabilities for surviving each infection combination $\prod_{r=1}^{r=n} p(s_r p(P_r))$, weighted by infection probabilities ($p(P_r)$) were then summed to provide the genotype fitness $(1 - s_H)$.

$$(1 - s_H) \leftarrow \sum_{c=1}^{N} \left(\prod_{r=1}^{n} p(s_r, p(P_r)) \right) \tag{2.3}$$

Initialisation of the model was performed over a range of settings in three parameters depending on the question. The three factors were: *i*) A factor influencing host mortality (parasite virulence, s_r from equation 3) over a range of 0 to 1, weakly to highly virulent parasite strains. *ii*) The effect of the recombination modifier, varying from no effect (complete linkage) to free recombination 0.0 to 0.5). *iii*) The degree of perturbance to the system (away from equilibria); this was done by increasing the skew of the initial parasite frequency distribution.

The results are shown using R (2.13.1, Figure 2.2 & 2.5) (R Core Team, 2013) and gnuplot (4.4, Figures 2.3 & 2.4) (Williams & Kelly, 2003), Figure 2.1 was produced using Dia (v0.97).

2.3 Results

Individual simulations

Figure 2.2 compares the RQH frequency oscillations in the diploid host with those in haplo-diploids. Whereas the predicted Red Queen oscillations can be seen in the diploid hosts in a vast range of the parameter space these are completely lacking in the haplo-diploids. To better understand this phenomenon we included extreme parameter values of the virulence, parasite range and recombination modifier.

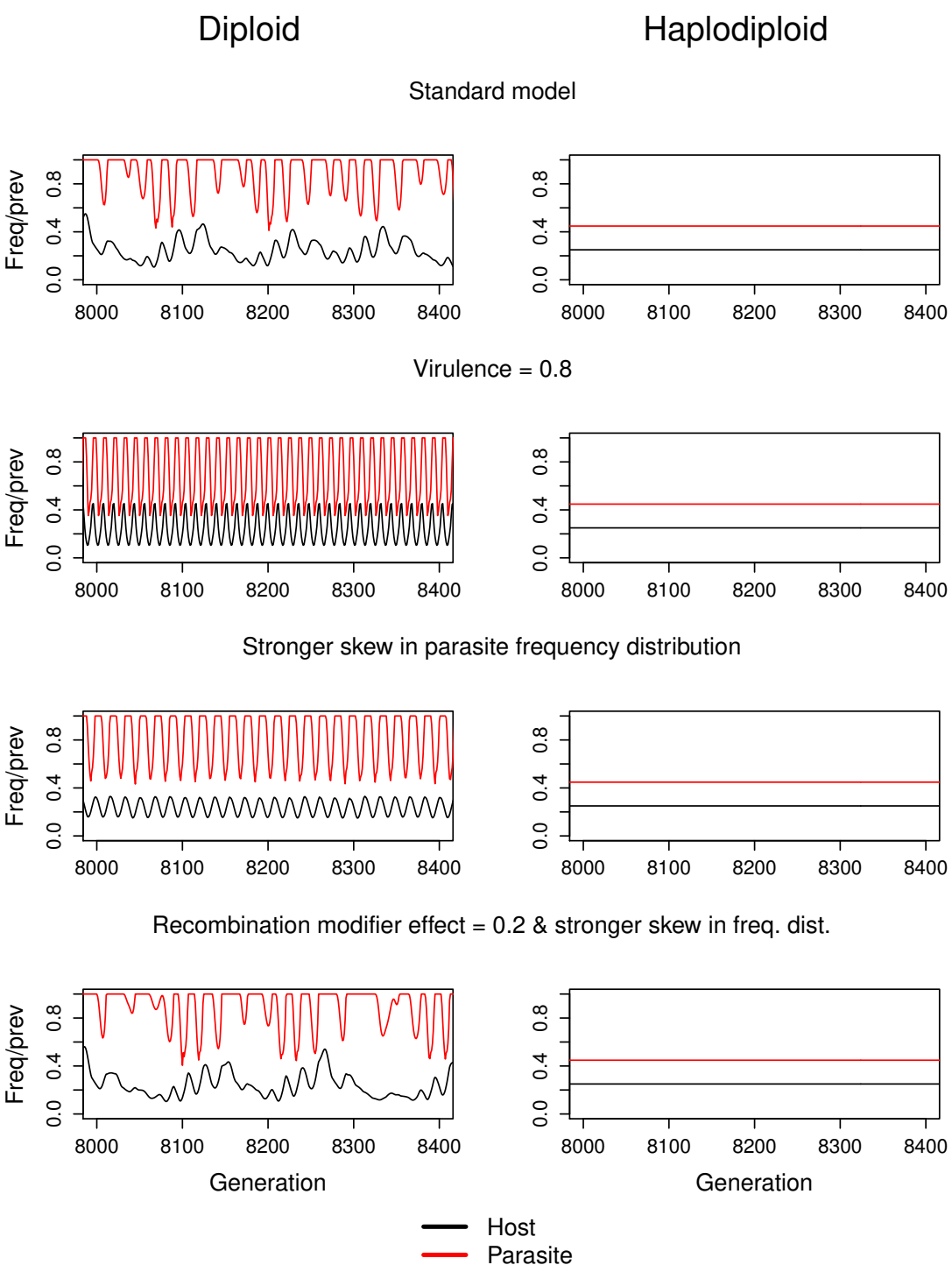

Figure 2.2: **Plots of the haplotype frequencies from diploid and haplodiploid simulations, and their respective parasite haplotype** (only 400 generations shown at steady state for clarity of presentation, 8000 to 8400 from 100,000 generations). Conditions in the standard run were as follows: virulence = 0.5, recombination modifier = 0.5, skewness of parasite frequencies = $3.6E^{-03}$. In the subsequent figures one or more parameters were changed (the stronger skew is upto $3E^{-02}$), these are shown above the corresponding figures. While oscillations in frequency were consistently observed in diploid host populations (over 100,000 generations), no such dynamics were observed with haplodiploid hosts. In the last graph for the diploid host population, it took upwards of 7000 generations until predictable oscillations occurred. Reduction of the modifier effect was done through the modifier allele M (as allele m has no effect), in the last two graphs M = 0.2.

Analyses of parameter space

Heatmaps illustrate the parameter space that facilitates the spread of the recombination modifier allele as predicted from the RQ hypothesis (Figure 2.3): That an arms race between parasites and their hosts will tend to select for elevated recombination rates in the host species. There are clear and pronounced qualitative difference between haplodiploid and diploid hosts regarding the frequency of the recombination modifier. As expected there is a broad parameter space favouring the spread of recombination in the diploid hosts (Figure 2.3, panel A). This parameter space is much smaller in the haplodiploid host population (panels B & D, Figure 2.3). Only under both an exceptionally high host mortality and a strong skew in the initial parasite frequency distribution will recombination between the two loci be favoured. The recombination modifier generally evolved to high frequencies when parasite viruence (s_R) was greater than 50%. The only exception was when parasite virulence was very high (> 0.95). This effect was driven by the extinction of the parasite strains early in the simulation, preventing further evolution of the recombination modifier. Selection in the haploids eliminated the matching parasite strains in vast regions of the parameter space.

A role of the siye of the effect of the recombination modifier allele (graphs C & D) an impact was only observed in diploid host populations in conjunction with low parasite virulence ($s_R < 0.5$). Low recombination modifier effects led to an increase of the Evolutionary Stable Strategy (ESS) frequency of the recombination modifier (panel C). In male-haploid populations the size of the recombination modifier has no impact on the range of parameter space where recombination is favoured. Surprisingly, with an increasing effect of the recombination modifier its frequency is reduced in the host population (panel D).

Fitness analysis

We explored the impact of virulence on host fitness by calculating the geometric mean fitness over the size of the recombination modifier effect and parasite virulence (Figure 2.4, graphs C & D). These results show a substantial decrease in host fitness over parasite virulence. There were no effects from the size of the recombination modifier effect in neither diploids, nor haplodiploids. The fitness of diploid host populations decreases much faster with increasing parasite virulence than in haplodiploids (graph C, of Figure 2.4). Under high levels of parasite virulence the haplodiploids can have an up to 2.6 fold advantage over diploid hosts (Figure 2.5).

2.4 Discussion

Individual simulations

We observed marked differences in the dynamics between the diploid and male-haploid host populations in response to infection by multiple parasite strains. The most striking result is the lack of the spread of a recombination modifier and the lack of Red Queen oscillations in haplodiploid

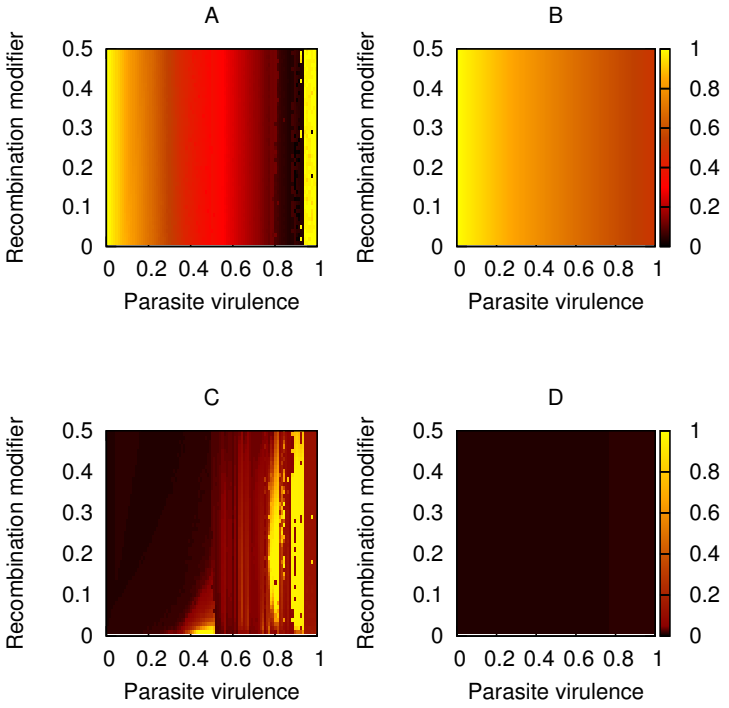

Figure 2.3: **Illustration of recombination modifier frequencies (z/colour-axis) in diploid (graphs in the left column; A, C) and haplodiploid (graphs in the right column; B, D) host populations.** With rates compared over various levels of parasite virulence (x-axis), different initial parasite frequency distributions (A & B) and sizes of the recombination modifier effect (C & D). In graphs A & B the recombination rate was fixed at 0.2, while the range of parasite frequencies was fixed to 0.8 in graphs C & D. Over most of parameter space only virulence can be seen to have an effect on the frequency of the recombination modifier (graphs A & D). However with a high initial range of parasite frequencies, an affect from the modifier can be seen in both diploid and haplodiploid populations. In diploid host populations lowering the modifier effect increases the range of parameter space in which higher modifier frequencies evolve; while in haplodiploids decreasing the effect size increases the frequency (over a limited parameter space) to which the modifier evolves.

host populations. Previous studies analyzed either invasion scenarios (M'Gonigle & Otto, 2011), advantages of the host to resistance (Oswald & Nuismer, 2007), or conditions favouring parasitism (Nuismer & Otto, 2004). The host systems of M'Gonigle & Otto (2011); Nuismer & Otto (2004) were either haploid, diploid (M'Gonigle & Otto, 2011); or diploid, tetraploid (Nuismer & Otto, 2004). In summary, these papers on population ploidy structure (Nuismer & Otto, 2004; Oswald & Nuismer, 2007; M'Gonigle & Otto, 2011) make the same general findings: *i*) haploidy (in the parasite) is almost a universally preferred state for parasitism; *ii*) in the MAM lower ploidy levels are preferred in the host species; *iii*) the opposite is true for the IMAM, here higher

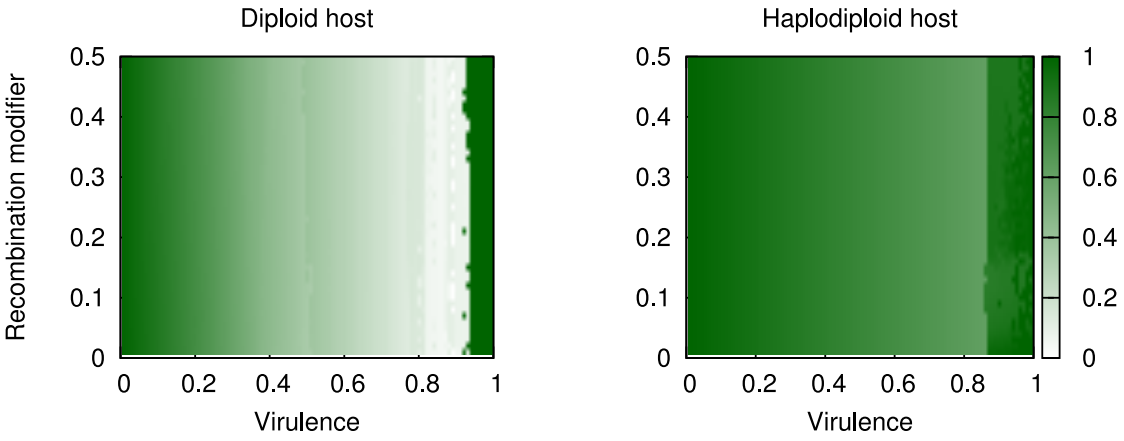

Figure 2.4: **Comparison of fitness over different recombination and virulence schemes, same to that of figure 2.3 graphs C & D.** The size of the recombination modifier effect has little impact on host fitness.

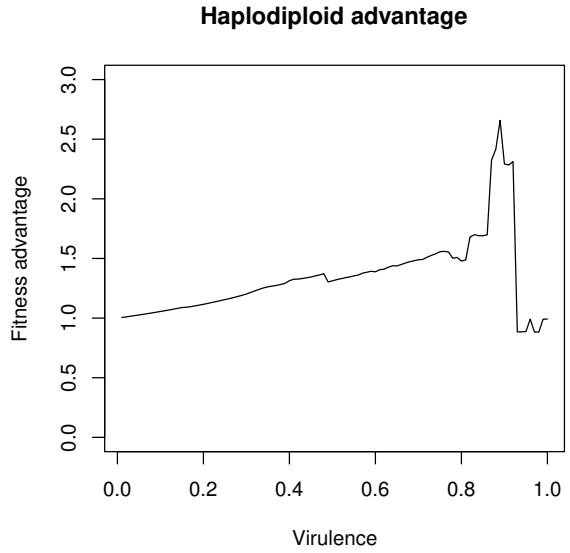

Figure 2.5: **The change in the x-fold fitness advantage of haplodiploid lineages against diploids, over virulence.** The drop in advantage at high virulence levels occurred from extinction of parasite strains, after successive parasite outbreaks. Only limited conditions show an advantage to diploid host populations.

ploidy levels tend to be preferred. In contrast, our data indicates that oscillations occur in neither the parasite nor host genotype frequencies in male-haploid host populations.

The two model systems presented here only differ in the genomic composition between the two host populations. Hence, any difference in the results must arise from some factor generated by the particular genomic composition. As the results are largely independent of the effect of the recombination modifier (Figure 2.3, D), the impact of recombination within a single gender can be rejected as the cause for an absence of RQ dynamics in the haplodiploid host. Remaining factors are differences in host fitness (between haploid and diploid hosts), and differences in

the susceptible host frequencies. Given that host fitness depends on the proportion of the host that is infected, and that parasite transmission depends on susceptible frequencies (appendix B, in the section "algebraic analysis of infection dynamics"), it must be the increased variance in susceptible frequencies (resulting from the haploid sex) that drove the differences between the two host populations.

Analyses of parameter space

The results regarding the recombination modifier in the diploid population were similar to those previously obtained by Agrawal (2009). Independent of the size of the modifier effect, higher recombination rates were generally favoured when virulence was high. In addition however, there is a small range in the virulence parameter space in the diploid host population ($0.45 \geq s_R \geq 0.5$) where the size of the effect of the recombination modifier could be small ($0.001 < \text{effect} < 0.02$) but nevertheless go to fixation. This does not occur in haplodiploids. Here the establishment of the modifier in the population is much less dependent on its effect size and also the initial parasite frequencies had little effect upon generating oscillations (Figure 2.3, graph A). Parasite virulence was the only driving factor, selecting for the spread of the recombination modifier.

The absence of both parasite and host genotype frequency oscillations in haplodiploids obviously removed recombination as an important driver of selection. This extended over a large region of parameter space and only extremely high virulence ($s_R \geq 0.85$) and highly disturbed situations (a large variance in initial parasite frequencies) generated selection for recombination in haplodiploids (Figure 2.3, graph B). Overall the smaller proportion of susceptible haploid male hosts limited the spread of parasite strains, reducing the strength of selection on the host. Further, whenever oscillations were sustained, the relationship between evolutionary stable strategy (ESS) recombination modifier frequencies and recombination modifier effects were negative (FIgure 2.3, graph D). This suggests that sex (or a high recombination rate) is not advantageous *per se* for haplodiploids in the context of host parasite evolution. Although our results by no means question the role of sexual recombination for selection in host-pathogen systems in diploid species, they indicate that Red Queen processes are less relevant for maintaining recombination (and sex) in haplodiploids.

Interestingly, extremely high recombination rates have been found in genomes of several eusocial haplodiploid Hymenoptera (Solignac et al., 2004; Meznar et al., 2010; Stolle et al., 2011), which is not predicted in our model. However, living in colonies dramatically changes the population structures for both hosts and parasites fundamentally violating many Hardy-Weinberg assumptions. For example there will be a dramatic increase of associative contacts among individuals with high genetic similarity. Moreover, selection at the colony level becomes important as does the reduced effective reproductive population sizes in the host. Selection on parasites will rely on both transmission within and between colonies, processes not addressed in our model. All these issues will clearly need to be considered when explaining the high recombination rates in eusocial insects (Solignac et al., 2004; Wilfert et al., 2007b; Meznar et al.,

2010; Stolle et al., 2011) These factors will also be important for any theory on parasite prevalence (van Baalen & Beekman, 2006; Hughes & Boomsma, 2006) and transmission (Walker & Hughes, 2009; Ottersatter & Thomson, 2007) in and across social haplodiploid species. In addition, it will be essential to understand the interactions between genetic diversity and sociality on parasite transmission (Baer & Schmid-Hempel, 1999; Hughes & Boomsma, 2006; Walker & Hughes, 2009; Ugelvig et al., 2010). For now, the robust empirical data on the impact of sociaility of parasites outside of human diseases is lacking (Craft & Caillaud, 2011).

Our results may have far reaching consequences not only for male-haploid, but also for diploid systems with hemizygous sex-chromosomes. Traditionally, a haploid genome (or hemizygous state) has been treated as a condition where individual alleles are under greater selective pressure (Smith, 2000). However, our results suggest that under the RQH, the proportion of susceptible individuals in the population is the dominant factor. This may explain the relative absence of selection observed on resistance loci on the X-chromosome (Hill-Burns & Clark, 2009), despite evidence for stronger selection over the entire X-chromosome compared to autosomes (Singh et al., 2008). Sex appears to be disadvantagious within male-haploid populations particularly if highly virulent pathogens are present. These results suggest that the high rates of recombination observed in the Hymenoptera may require alternative explanations than that provided by the RQH.

2.5 Conclusion

We observed a marked difference in the host-parasite MAM dynamics between male-haploid and diploid host populations. These differences in host population dynamics have a strong impact on fitness, and the evolution of recombination modifiers. This results in fundamental differences in the predictions of the RQH regarding male-haploid host populations to parasitism, compared to the traditionally studied diploid host populations. Our results suggest that the evolution of high recombination in social Hymenoptera may require an explanation alternative to the RQH.

Acknowledgements

We would like to thank the Deutsche Forschungsgemeinschaft (DFG) for fundign provided as part of the SPP Project number 1399. We would also like to thank Fouks B., Murray T and numerous reviewers for checking the manuscript.

Chapter 3

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

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 halpoid species.We here use Agent Based Models (ABMs) 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 MAM). 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.

3.1 Introduction

Evolutionary theories explaining the development and maintenance of sex, those involving meiosis, frequently employ the Red Queen Hypothesis (RQH). The Red Queen Hypothesis is derived from 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, 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 Salathe 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 & Lively, 2002; Nuismer & Otto, 2004; Otto & Nuismer, 2004; Peters & Lively, 1999, 2007; Schmid-Hempel & Jokela, 2002). Those studies that compared the ploidy of hosts on the RQH are few and far between (Nuismer & Otto, 2004; M'Gonigle & Otto, 2011; Kidner & Moritz, 2013) with only Nuismer & 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 & 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 & Lively, 2002; Agrawal, 2009; Hodgeson & Otto, 2012; M'Gonigle & Otto, 2011; Peters & 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 & Moritz, 2013; Lively, 2010). However, several restrictions of the mean field model have been recorded (Boots & 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 & 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., 2007a). 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 (Ottersatter & 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).

3.2 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 appendix C). 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

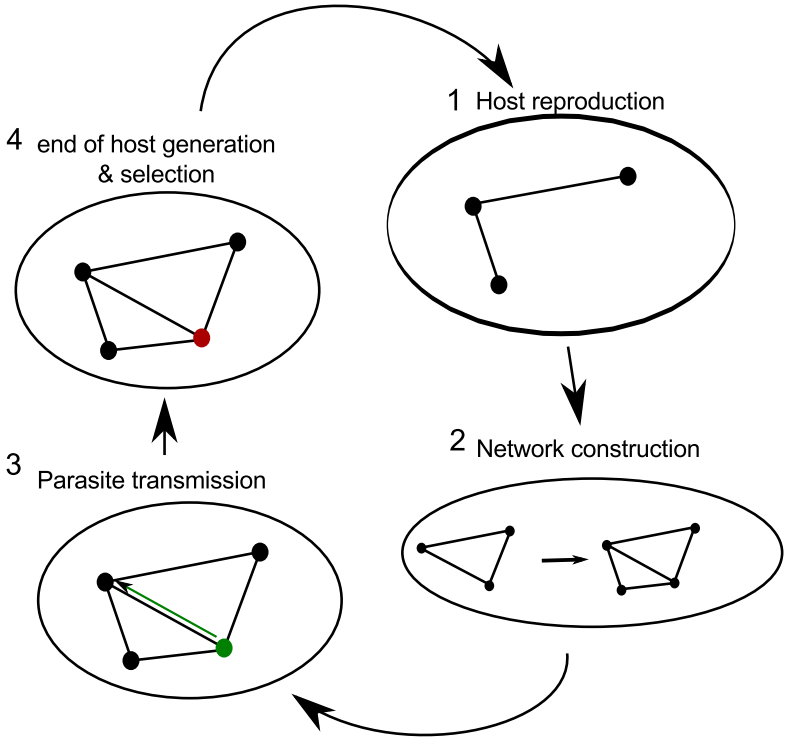

Figure 3.1: **The basic model layout, split into four parts**. *i)* The first required stage is to set up the host population. This is done by simulating host reproduction with a uniform systematic selection of host genotypes. *ii)* Afterwards, the host network is constructed using a scale-free algorithm for adding nodes. After this the network is then refined to meet the specific network parameters (cliqueness and connectedness). *iii)* 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. *iv)* 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 *i)* is entered.

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, appendix C: algorithms) the host network was constructed (algorithms 1 – 3, appendix C: algorithms). Inclusion of the first and second nodes requires the addition of a single edge between these two 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 (Figure 3.1 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 individual‘s 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). 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 1000 host generations before extinction of a susceptible population.

Analyses

The population size was set to 4000 and the number of host generations to 5000. These values were used to reduce the impact of stochasticity on the host, while also reducing computational time. 5000 host generations was chosen after running of the simulation (after 3000 to 4500 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 (initialy 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 & 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)).

3.3 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 (Figure 3.2). This has two major consequences: Firstly, the overall frequency of susceptible hosts is higher in the diploid population (from an average of 0.148 to 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 Figure 3.2) 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 Figure 3.2). 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 4000 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 appendix C: table 1) 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 & Lively, 2007).

Host fitness

The relationship between fitness of the host and interaction model follows that of parasite prevalence (Figure 3.3). 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)). 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 (Figure 3.4). 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. In contrast, the connectedness (or edge density) has a distinctly non-linear relationship when I-S edges are abundant.

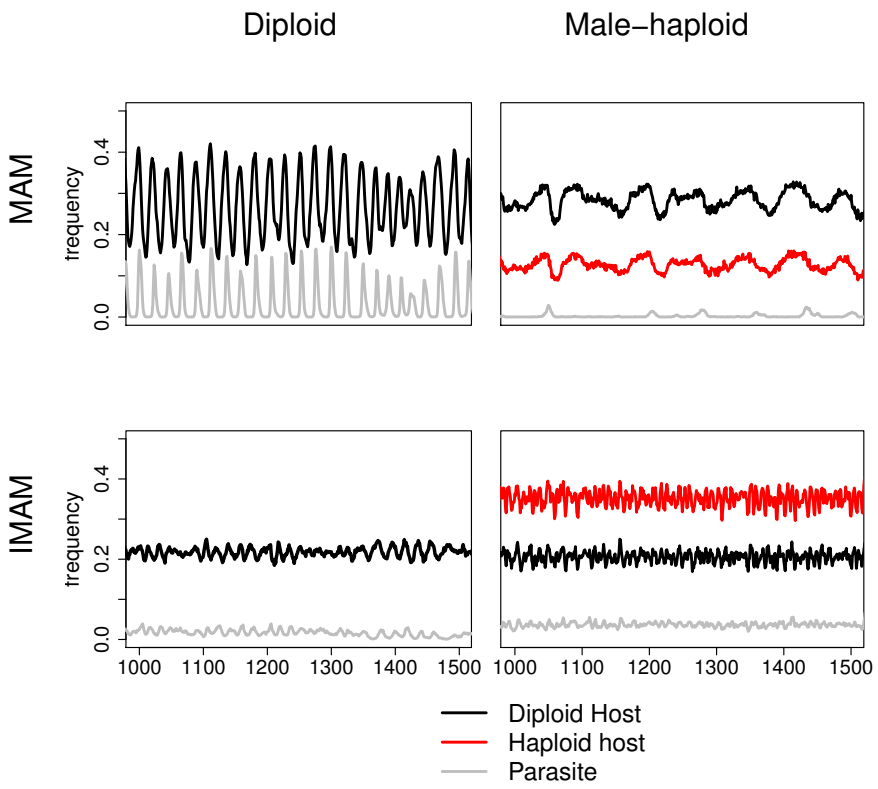

Figure 3.2: **The change in frequencies of the parasite strain *ab* and the corresponding proportion of the host susceptible to this parasite strain.** The host frequencies provided are based on the gender to illustrate the impact of ploidy on the frequency of susceptible hosts within the population. 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.

Interactions between the model parameters are less evident (Figures 3.5 & 3.6). In general host fitness is highest when the parasite-host generation ratios are low and virulence is high (Figure 3.5). 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 (Figure 3.6).

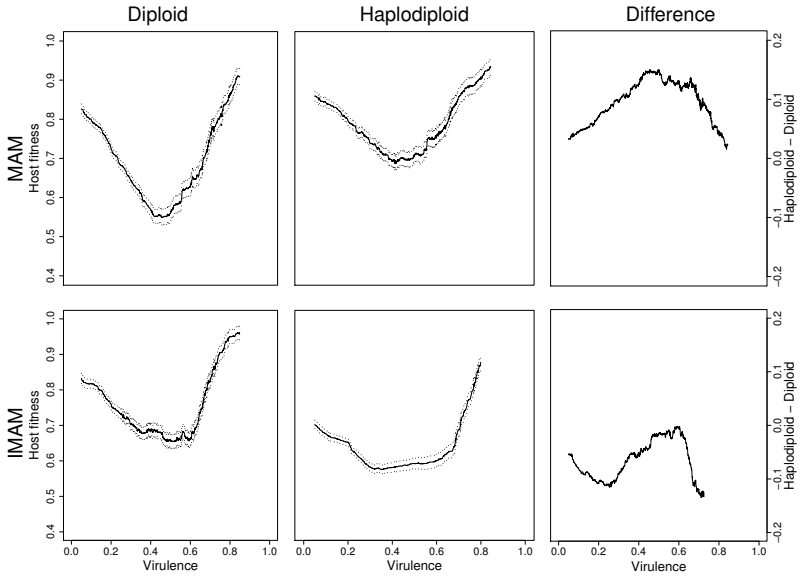

Figure 3.3: **The relationship between parasite virulence and host fitness.** Fitness is calculated systematically throughout the simulation up to 200 times (except when the simulations stopped prematurely from the extinction of multiple host haplotypes). The median and the 95% confidence interval (dotted 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

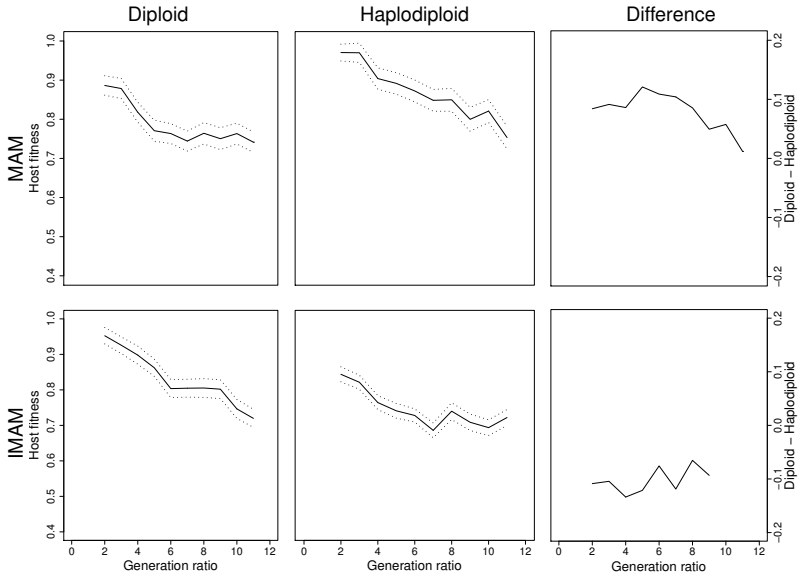

Figure 3.4: **The relationship of the host - parasite generation ratio and host population fitness.** Fitness is calculated as in the previous figure using a sliding window, the 95% confidence interval is shown by the dotted lines. The lower fitness observed for diploids under the MAM (and haplodiploids 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.

including cliqueness only the interaction term with parasite virulence was retained (all other terms were retained).

3.4 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 (Figure 3.2). 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 all three the overall number of infected hosts, the prevalence of multiple infections, and the parasite virulence (Figure 3.5) as previously reported by Roode et al. (2008): Diploid hosts have lower fitness values over a range of parasite virulence under the MAM, higher fitnesses under the IMAM (Nuismer & Otto, 2004; Oswald & Nuismer, 2007; M'Gonigle & 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 (Figure 3.3). 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, $W = 540205$, 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 (Figure 3.5). A result similar to those from previous studies (King et al., 2012), and in support of the invasion scenarios of M'Gonigle & Otto (2011). The results here being attributable to the proportion of I-S edges, and the general density of edges in the network. In figure 3.6 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 & Otto, 2006; Nuismer & Otto, 2004; Otto & Nuismer, 2004; Peters & Lively, 1999, 2007; Schmid-Hempel & 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; Salathe et al., 2007). Secondly, high transmission rates are generally required for selection on recombination. 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., 2007a; Meznar et al., 2010; Stolle et al., 2011).

Network vs mean field vs single challenge models

Two general factors are important when considering the type of model to use: i) model complexity, ii) analytical power. While mean field models are both simpler and more powerful analyticaly, 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: i) the significance of the cliqueness by virulence interaction term, ii) 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'.

acknowledgements

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).

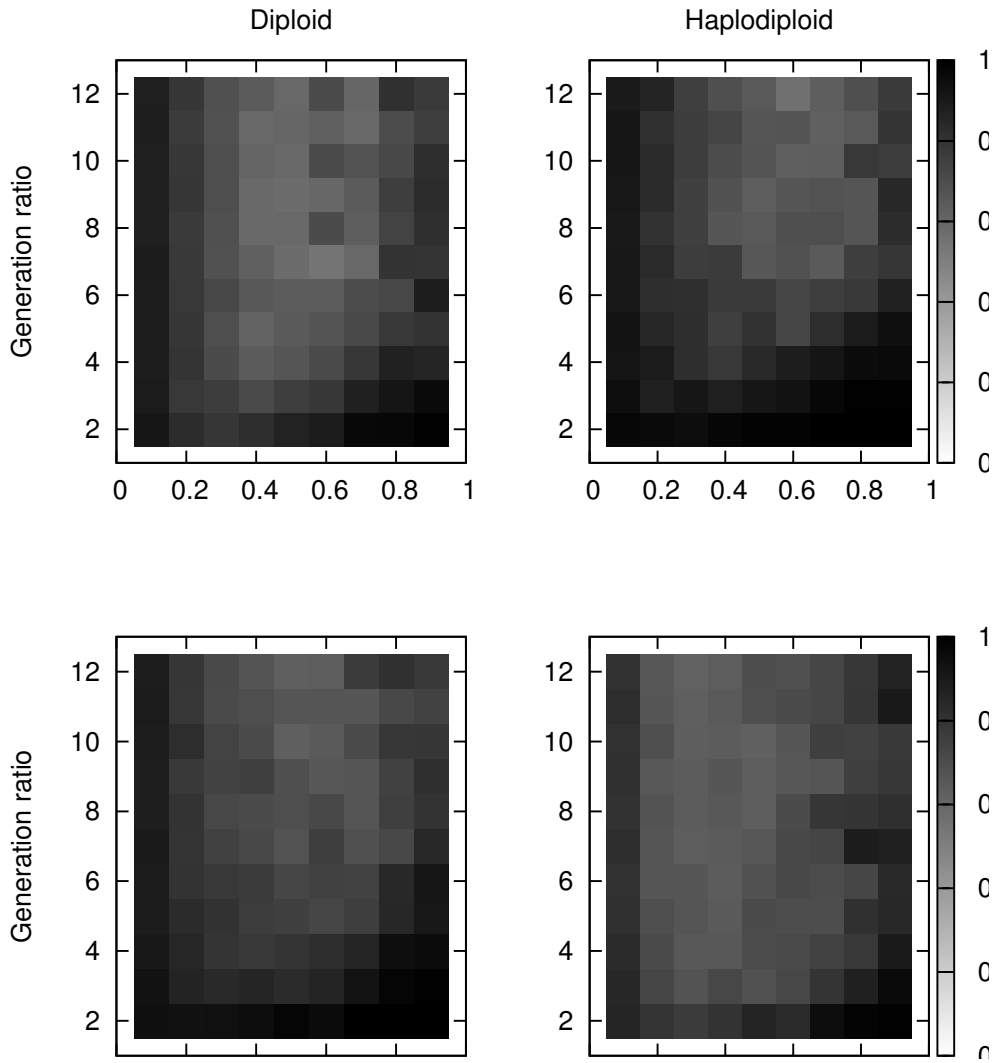

Figure 3.5: **A heatmap showing the distribution of host fitness depending on the host reproductive system (Diploid, or Haplodiploid) and interaction model (MAM, above; IMAM below).** The heatmaps are plotted over a range of transmission values (host-parasite generation ratio) and virulence (probability 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.

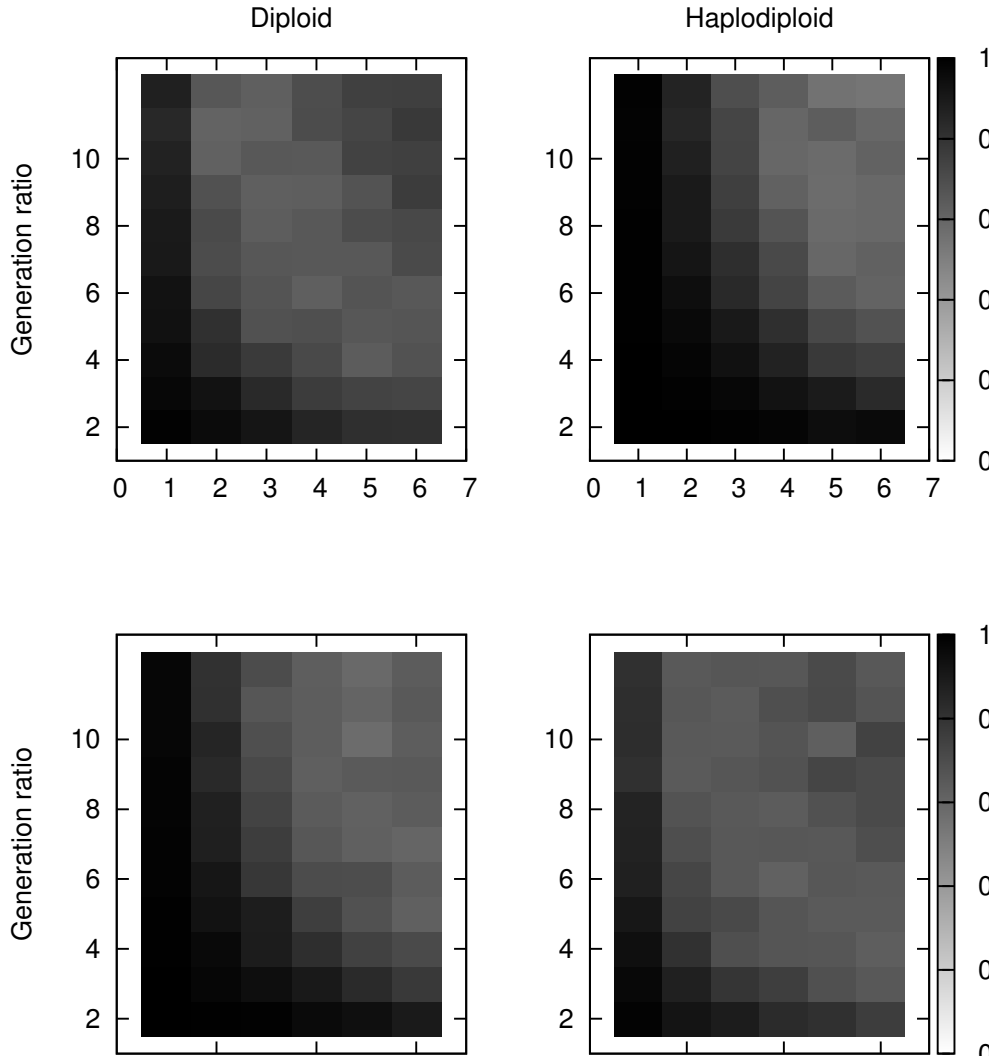

Figure 3.6: **A comparison using heatmaps of the host fitness depending on reproductive system and interaction model (MAM, above; IMAM, below).** The x-axis here uses the average number of connections simulated between host individuals. A strong interaction between the transmission parameters is clear 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).

Part IV

Conclusions

Chapter 4

Synthesis

4.1 Part II

In chapter 1, I looked at the potential for the MTH (Maternal Transmission Hypothesis) and DMH (Deleterious Mutation Hypothesis) to explain the evolution of male-haploidy in ancestrally diploid populations. From this it was possible to identify the range of parameters that are conducive to the invasion of male-haploidy in a population (assortative mating > 0, the haploid viability cost ≤ 0.5). Upon introducing stochasticity into the simulations the selection coefficient and degree of heterozygosity were found to no longer provide limiting conditions for the evolution of male-haploidy (compare Figure 1.4 to 1.2). In general, evolutionary models have avoided the inclusion of stochastic noise/variation due to computational limitations (in memory and computing time). With the progression of computing power (according to Moore's law), however, stochastic models/simulations are becoming increasingly feasible and legitimate as a tool for studying evolution. Accordingly a key point in chapter 1 is the comparison between stochastic and deterministic simulations; the former showing larger regions of parameter space conducive to the invasion models. A further conclusion (using MTH corrections for the genotype frequencies) was the dominant role that the MTH plays in explaining the results. I note, though, that its predominance could be exaggerated through the use of a single locus viability model.

While the model of chapter 1 is derived from embryonic PGE (Paternal Genome Elimination), my model in no way encapsulates the body of processes present in this system; the model presented reflects the means by which genetic mutations may be differently expressed between this and diploidy. Other processes that can be reasonably expected to play a broad role in the evolution of male-haploidy are: *i*) Sexually antagonistic selection, competing interests between the sexes can lead to the development of an Evolutionarily Stable Strategy, to the detriment of one sex. *ii*) Cytoplasmic incompatibility (CI, including maternally inherited symbionts, or other maternally inherited elements), elements that are cytoplasmically inherited have an interest in killing males as they present an evolutionary dead end. This theory relies on the ability of the

host (of the cytoplasmic elements) to resist the male killing effects of the cytoplasmic agents. *iii*) Sex ratio selection, this concept involves measures of inclusive fitness and competition over sex allocation rates (Hamilton, 1967). *iv*) Intra-genomic conflict arising as a result of parental care of offspring and conflicts between paternal and maternal investment. In most case studies, two or more hypotheses are likely to be involved; in the Sciaridae, genomic imprinting (Goday & Esteban, 2001) (arising through intra-genomic conflicts) and drive[1] (the MTH) are strongly expected to play a role in the evolution of male-haploidy. The DMH is less likely to play a role in the Sciaridae due to the presence of germline PGE (paternal chromosomes are inactivated only in germline cells) (Ross et al., 2011). For the Aleyrodidae, CI (we conflate maternally inherited endosymbionts with CI here, due to the similar transmission mechanisms and impacts) is a highly probable causal factor due to maternally inherited bacterial endosymbionts (Ross et al., 2010) that are crucial for the digestion of phloem sap. The role of CI is also likely to be high in those cases where *Wolbachia* has been identified as a driver (Ross et al., 2012b).

The focus in chapter 1 on the MTH and DMH was due to these two theories providing an inherent component of an invasion model for male-haploidy. The MTH is a consequence of drive, a product of the unbalanced inheritance mechanisms in male-haploid groups. This process applies to all cases of male-haploidy, as the paternal genomes in all instances are not heritable through the male-lines, thus providing an inherent advantage to maternally inherited elements in males. Hence any consideration of the occurrence of male-haploidy should likely take this hypothesis into account. An issue regarding the case for the DMH in previous studies has been the usage of single locus viability models (Goldstein, 1994), and/or the usage of male-viability cost models (Smith, 2000). Single locus viability models describe the determination of a hosts fitness through mutation at a single locus. Chromosomal viability models, where mutation across the loci on a chromosome determines the host fitness, in contrast remains absent from the literature on the evolution of male-haploidy. It is entirely possible that previous results on the DMH will not hold under either chromosomal, or genome-wide viability models. Such models should be possible as estimates are available for genome-wide mutation rates (Eyre-Walker & Keightley, 1999; Nachman & Crowell, 2000) and for the distribution of mutational effects (Loewe & Charlesworth, 2006). A further point in regard to the role of mutation (and the DMH) arises through the incorporation of stochastic simulation models (as in Chapter 1 here). While the incorporation of stochasticity has little impact on fixed viability costs (parameter c in Figure 1.3), it has a very large impact on the properties of the viability locus (Figure 1.4, where mutation applies). The variance arising from mutation in multi-locus models (genome/chromosome viability models) may be essential in resolving the role of the DMH in male-haploidy.

Given current computational resources, it is entirely possible to generate an agent based genome viability model for investigating the potential for male-haploidy to invade. This may be done through the development of a Muller's ratchet model for individual chromosomes. Muller's ratchet describes the process by which deleterious mutations can accumulate in populations

[1]The MTH can be considered as a form of genetic drive, the genetic agent causing the MTH provides an advantage that can overcome moderate selective forces (a haploid viability cost of 0.5).

through purely random processes (Haigh, 1978). Deleterious mutations are assumed to accumulate in the population through repeated mutational events and the loss of fitter genotypes through genetic drift. Such a model can be applied to each separate chromosome (the chromosomes acting as an agent), with the product of the chromosomal fitnesses used to calculate individual level fitnesses. This system subsequently generates the genome-viability model, cursory results of which, on the invasion of male-haploidy, can be observed in Figure 4.1.

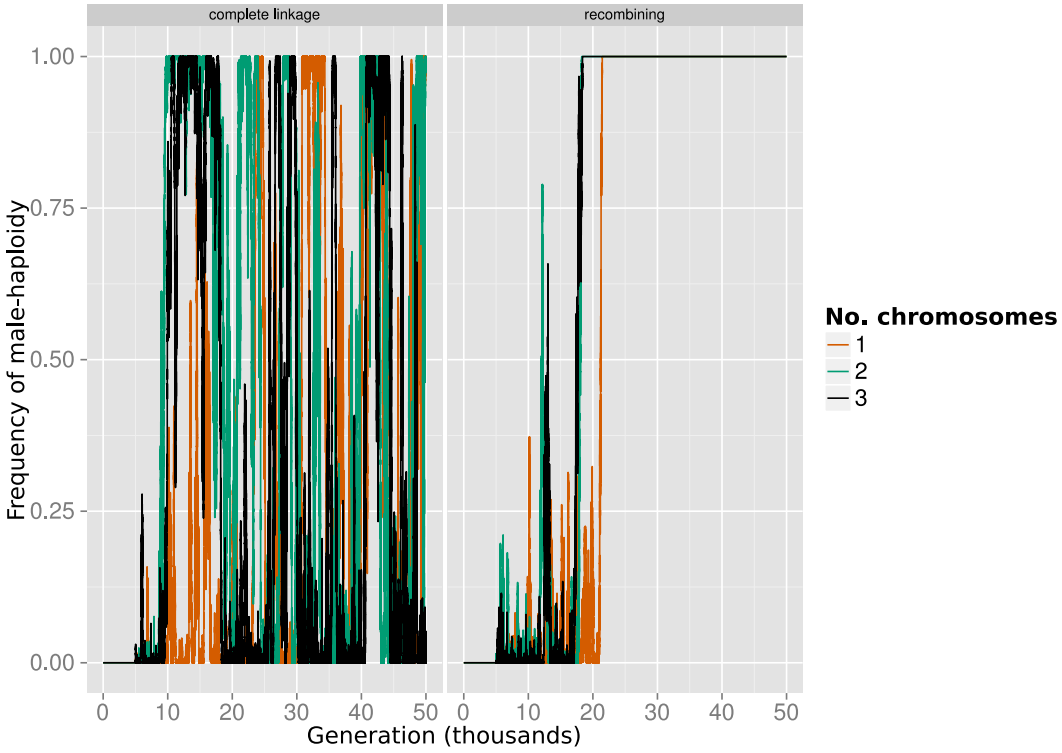

Figure 4.1: **An illustration of preliminary results from the use of Mullers ratchet to simulate a genome-viability model for the invasion of male-haploidy**. In the Mullers ratchet model developed, individual loci can be considered as the lowest level at which the agent based model may operate. As such although incorporation of recombination traditionally violates the assumptions of Mullers ratchet, this is not the case with the agent based model developed here. The main conclusion observable in these results is that recombination is required for the stable invasion of male-haploidy, under a genome-based viability model. In its absence diploidy is able to re-establish itself, after the invasion of male-haploidy.

Part II summary

While the power of the DMH is not currently ascertainable, due to the dependence on single locus models, the power of the MTH is. Consequently the results, regarding invasions of diploid species, from single locus models show the dominance of the MTH. While the DMH cannot

explain the ability to overcome large male-haploid costs, the MTH can. The introduction of stochastic variation largely expanded the parameter range for invasion. This observation could be the result of probabilistic models allowing a random walk through the genotype frequency/fitness landscape. Unlike in deterministic models, probabilistic models allow for variation in model variables (in this instance the genotype frequencies); while frequencies are not independent between discrete generations, they are considerably less dependent than in deterministic models. With two competing forces (heterozygote advantage promoting diversity at the modifier locus, the MTH promoting the invasion of male-haploidy) a probabilistic model is likely to favour direct over indirect forces, particularly if the direct force is stronger (as the MTH is here). Hence the observations made in Chapter 1.

Genome-viability models can be easily adopted in agent based models to study the invasion dynamics of male-haploidy. Utilizing these models demonstrate that male-haploidy may invade diploid populations, but the stability of invasion is dependent on recombination. These models can also provide a platform for estimating the relative strengths of the DMH and MTH, however estimates on the DMH are limited by the current availability of data (Charlesworth & Charlesworth, 1998; Eyre-Walker & Keightley, 1999; Nachman & Crowell, 2000; Loewe & Charlesworth, 2006).

4.2 Part III

On the Red Queen

Parasite transmission was excluded in the early Red Queen literature (Bull, 1979). The system used instead was the "single challenge model", in which every host is "challenged" by a single parasite. The use of this system, compared to parasite transmission, allows a tighter coupling between host and parasite demographics. The "single challenge" systems also facilitated the use of analytical approaches for modeling the Red Queen. Use of this model enabled Bull (1979), and later others (Barton, 1995), to ascertain that host-parasite coevolution could explain the evolution of sex and recombination. More recent studies (Salathe et al., 2008; Agrawal, 2009) continue to follow this approach, applying further features, two examples being: *i*) differences in generation times between parasites and hosts (Salathe et al., 2007), and *ii*) discerning the different selection regimes for sex and recombination. Fewer studies have focused on parasite-mediated factors (Craft & Caillaud, 2011; Lively, 2010), despite acknowledgement of their role in host-parasite coevolution (Lively, 2010) and the RQH. This is despite the many implicit relationships between host-parasite interactions and parasite transmission. Some examples of this are:

i) the genetic interaction model,

ii) the fitness function, and

iii) populations genetic structure.

These implicit relationships can make models analytically demanding; computational approaches do, however, remain a possibility.

Chapter 2

In Chapter 2, I introduced a simplistic, population genetic simulation. In this model a very basic parasite transmission model was adopted in which infected individuals provide a source for future infection events (failure to infect is possible). This differs from the earlier "single challenge" system by allowing parasites to "under infect" the host population. Incorporating this feature (Kidner & Moritz, 2013) led to results that were divergent to those that were "expected" (Schmid-Hempel & Jokela, 2002), noting that Schmid-Hempel & Jokela (2002) attempted to model a similar system but with a social context and without male-haploidy. It should also be noted that, while only a simple parasite model was used, it still required the following assumptions: *i*) parasite strains/species infect independently of each other, and *ii*) host fitness functions from multiple infections were multiplicative and independent. Incorporating these features led to the breakdown of the host-parasite frequency oscillations in haplo-diploid hosts (Kidner & Moritz, 2013), while these oscillations remained in the distinct diploid host populations. Two factors could underlie this breakdown of the host-parasite frequency oscillations: *i*) transmission effects - "herd immunity", and/or *ii*) the translation of infection to host fitness. Isolating these two effects while changing host ploidy (from diploid-diploid to haploid-diploid) proved to be too difficult to implement under this model (Kidner & Moritz, 2013). Discerning these different effects proved easier with a more complete picture of parasite demographics (including meaningful parameters that could be varied) (Kidner & Moritz, 2015).

Chapter 3

In Chapter 3 I introduced a host-parasite coevolutionary model that allowed me to analyse parasite transmission and host fitness separately. In contrast to the use of a mean field model in the study of Lively (2010), I employed a full network model. The parasite's environment in the full network model was described by the average number of neighbours, and the degree to which these neighbours were interconnected (edge density, and cliqueness, respectively). (Cliques describe the way neighbourhoods form in a network; in technical terms a clique is a subset of nodes in a network that predominantly share edges among themselves). Varying these parameters enabled a range of parasite demographics, allowing us to test the RQH under a variety of scenarios. Under the previously made assumption that "herd immunity" is one of the major processes determining the previous simulation results (from Chapter 2), two parameters would be expected to have an inordinate role in producing the observed results: *i*) the number of infected and susceptible nodes (meaning the host-parasite interaction model), and *ii*) the average density of edges per node.[2] The rate of transmission of infected - susceptible edges would not be expected to play

[2]In network terminology an edge describes a connection or interaction between two nodes. In this case the node is an individual in either a diploid or male-haploid population, hence the edge becomes a connection through

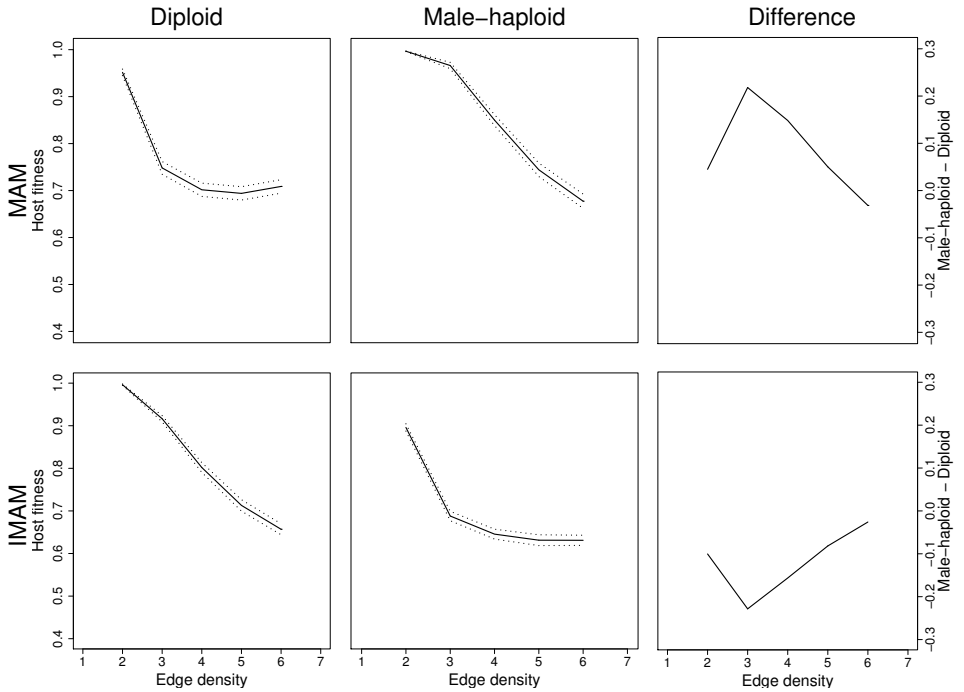

Figure 4.2: **A supplementary figure from the third chapter, showing the relationship between host fitness and the average density of edges per node.** These results show stronger responses with lower values of connectedness when there is a large proportion of susceptible nodes within the host population: Larger proportions of susceptible nodes are present in diploid populations under the MAM, in haplodiploid, or male-haploid populations, this occurs under the IMAM. The non-linearity in these cases represents the fact that under these conditions saturation of susceptible nodes (to the infected status) is rapid and will reach an asymptote.

a dominant role, as the effectiveness of this rate depends on the frequency of these edges. The same applies to the generation ratio, in addition to further confounding interactions with the clique parameter. If "herd immunity" caused the results in the earlier study (Kidner & Moritz, 2013), then different responses would be expected in host fitness (and parasite prevalence) to changing transmission between hosts (Figure 3.4). In contrast, if differences in virulence were the cause, then strong diverging responses should be expected from changing virulence (Figure 3.3). The addition of the inverse interaction model in Chapter 3 (Inverse Matching Allele Model, IMAM) enabled further comparisons regarding the response of host fitness to virulence and transmission. The approximately linear responses of fitness in diploids and male-haploids, under the IMAM and MAM respectively (Figure 4.2), indicate that the density of edges is important in determining host fitness (Figure 4.2). Comparing this with parasite demographics (whether

which parasites are transmittable.

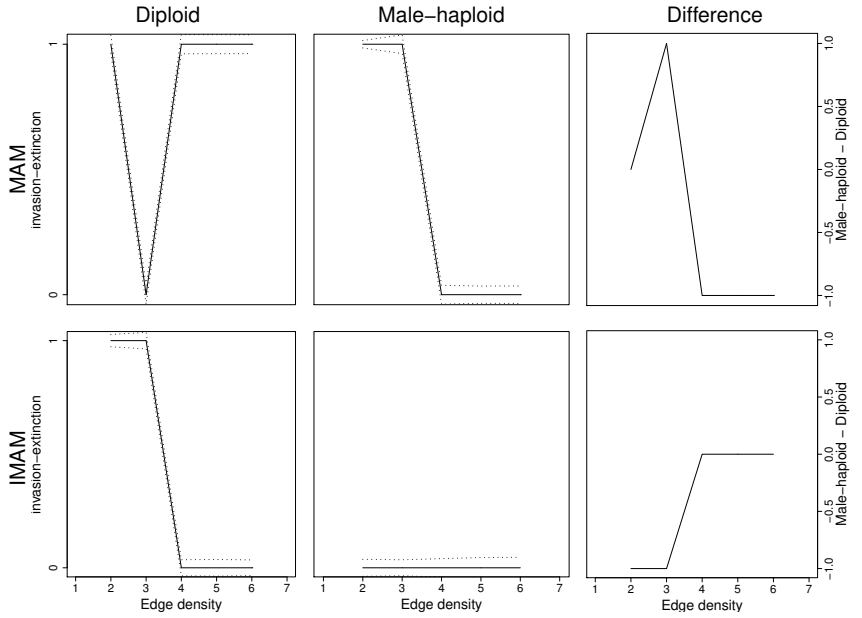

Figure 4.3: **A previously unpublished figure illustrating the type of population demographics existing in the parasite populations under different host scenarios.** The two values on the y-axis represent the type of demographic processes occuring in the parasite population: The value 1 indicates that the parasites are subject to a series of invasion extinction events, 0 relates to the situation when the parasites are able to maintain relatively stable populations in the host species. If the proportion of susceptible individuals is suspected to be a limiting factor (top right and bottom left), similar responses are observed with sustainable populations occurring when each node has an average of 4 edges. Diploid populations under the MAM are in contrast predicted to rarely lead to sustainable parasite populations; this is in stark contrast to male-haploid populations under the IMAM, which is not predicted to lead to parasite invasion-extinction dynamics.

parasites are subject to regular host-extinction cycles, or have sustainable populations) shows equivalent responses in these same scenarios (Figure 4.3). When susceptibility to parasites is a limiting factor, increasing the density of edges (to 4 edges per node) leads the parasites to sustainable population growth patterns.

In contrast, under the alternatives (when the degree of susceptibility is not a limiting factor), the differences under the IMAM and MAM are stark. Diploids under the MAM show little tendency towards sustainable parasite populations (the top left graph in Figure 4.3); while male-haploids, in contrast, show consistently sustainable populations under the IMAM (Figure 4.3, lower middle graph). These differences might arise from the distribution of susceptible individuals among the host populations: under the MAM, heterozygotes are susceptible to multiple parasite strains, while under the IMAM this occurs only in homozygotes. If a parasite invades a diploid

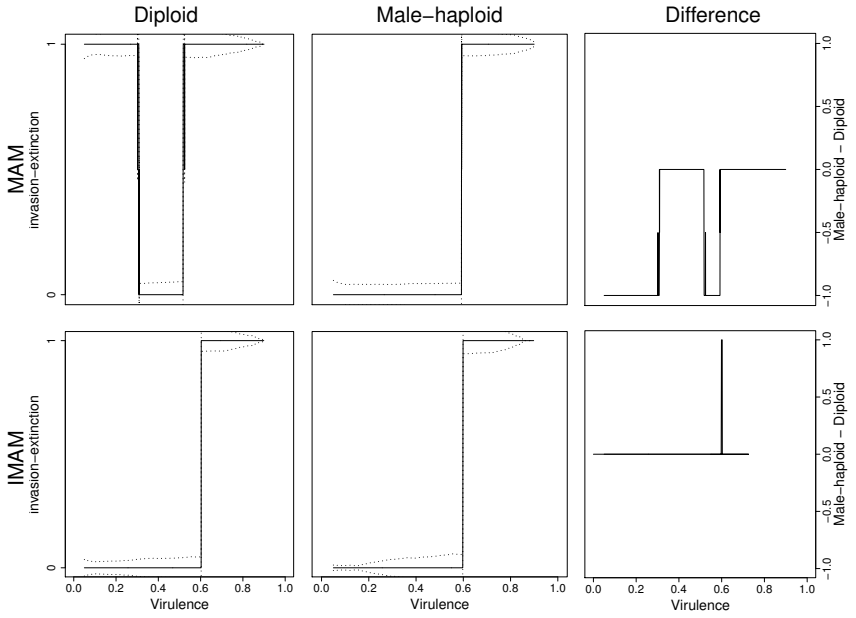

Figure 4.4: **An unpublished figure from chapter 3 showing switches between parasite population dynamics over changing parasite virulence.** Parasites under all conditions (except diploids under the MAM) show the same response, invasion-extinction dynamics dominant if virulence is greater than $^{3}/_{5}{}^{\text{ths}}$. In diploids invasion-extinction dynamics dominant above $^{3}/_{5}$ and below $^{2}/_{5}$, the remainder of virulence parameter space leads to stable parasite dynamics.

population, under MAM dynamics, then the susceptible haplotypes will decrease in frequency, leading to infections in predominantly heterozygous individuals. (With four haplotypes, the probability that a susceptible individual is homozygous at both loci is much lower; additionally, heterozygous individuals may also be infected by other parasites, both aspects could underlie the diploid results in the top left graph in Figure 4.4). Under invasion-extinction dynamics, the parasite strain will deplete the host of susceptible genotypes, at which point new infection events become too rare to support population growth or maintenance. In contrast, the main source of susceptible individuals within male-haploid hosts, under the IMAM, for any particular parasite strain are homozygotes, and may also be subject to infection from two other parasite strains. (Double heterozygotes under the IMAM are immune to infection by all parasite strains). Under these scenarios, host responses can be expected to generally result from multiple infections, with parasite interactions potentially playing a role.

If parasite transmission can play such a role in these RQH models, then why is it so often absent? A brief purview of the literature would suggest the reason is 'ease of analysis'; *i*) network-models tend to be complex, mean-field models (Keeling, 1999) less so, and *ii*) "single challenge" the simplest analytical approach. This does not, however, reflect the advances that have been

made with infection models in ecological epidemiology(Keeling, 1999; Haraguchi & Sasaki, 2000; Boots & Mealor, 2007; Ames et al., 2011; Webb et al., 2013). Nor does it reflect recognition in experimental studies of the roles of virulence and transmission (Roode et al., 2008). In regards to host-parasite coevolutionary studies, transmission is here considered to be as important as virulence.

Part III summary

The introduction of even a basic transmission model (under otherwise identical conditions) qualitatively changes the outcome of RQH simulations. The form introduced in chapter 2 showed that the RQH predictions are unstable over changes in host ploidy; whether this instability was a result of differences in fitness, or transmission could not be ascertained. Consequently, a network model in chapter 3 was developed to allow for a greater exploration of the role that transmission could play. In this network model the relationship between connectedness (average number of edges/connections between nodes/hosts) and fitness was observed to follow different functions based on the host and interaction model (Figure 4.2). These differences followed changes in the frequency of susceptible individuals. With lower degrees of susceptibility in the host populations more or less linear responses were observable (Figure 4.2). For parasite populations lower host susceptibility appears to be correlated with switches to invasion-extinction dynamics when each host has an average of four or more connections (Figure 4.3). In contrast when the degree of susceptibility is expected to be less limiting (Diploids under the MAM and male-haploids under the IMAM) parasite demographics diverge massively between systems. Parasites in male-haploid hosts under the IMAM appear to never enter invasion-extinction dynamics in almost complete contrast to parasites of diploids under the MAM. While this was also seen in relationships between parasite demographics and virulence in diploids (Figure 4.4), this was not the case for male-haploids. Diploids under the MAM tend to favour the invasion-extinction parasite dynamics over sustainable populations ($0 < virulence <^{1}/_{3}$ and $virulence > {}^{3}/_{5}$). In all other contexts (Figure 4.4) invasion-extinction dynamics occurred as virulence was greater than ${}^{3}/_{5}$. While these results do not demonstrate, with complete clarity, the role that transmission and virulence played in the results from chapter 2, they do demonstrate the need to accomodate transmission in future research on the RQH.

Results on recombination; how clear are the results?

Empirical observations of recombination rates in the Hymenoptera (Solignac et al., 2004; Wilfert et al., 2007a; Meznar et al., 2010; Stolle et al., 2011; Liu et al., 2015; Wallberg et al., 2015) all appear to show high recombination rates, the findings from the RQH (Chapters 2 and 3) were not able to account for this. Though recombination can plastically increase as a response to infection by parasites in *Drosophila melanogaster* (Singh et al., 2015). This can hardly prove surprising as many factors (sociality, skewed sex-ratios, parasite competition; for a few) were

still not included in those analyses, even with the addition of transmission models (Kidner & Moritz, 2015).

In the literature on the RQH from the 2000's Barton (1995) identified some of the conditions required for the evolution of recombination. Later literature more closely modelled the dynamics between hosts and parasites, from which came the recognition that Barton's conditions were too restrictive (Peters & Lively, 1999; Salathe et al., 2007; Peters & Lively, 2007). These models also allowed for the integration of other processes: sociality (Schmid-Hempel & Jokela, 2002), directional selection (Hodgeson & Otto, 2012), ploidy/sex-system (M'Gonigle & Otto, 2011) and mutation (Howard & Lively, 1994; West et al., 1999). One of the issues that has been previously discussed regarding the state of the Red Queen literature is the need to delve more deeply into factors implicit in the RQH (Salathe et al., 2008). Among those cited is multiple parasite infections, and dropping the consideration of infection as a random process. Even so, too few papers have considered parasite transmission (Lively, 2010), despite further calls for its inclusion (Craft & Caillaud, 2011).

While the evolution of recombination still requires substantial investigation, one current emphasis is on the use of pluralistic models (West et al., 1999). One possibility is the incorporation of mutation models (as in the use of models like in Chapter 1) with the RQH (Chapters 2 & 3). The dependence upon recombination from preliminary results on Muller's ratchet and invasion models of male-haploidy has already proved an interesting avenue of research (Howard & Lively, 1994). Given the research presented here in this thesis, it appears highly unlikely that the RQH alone can adequately explain experimental observations on recombination (Solignac et al., 2004; Meznar et al., 2010; Wallberg et al., 2015).

4.3 Final Comments

Chapters 2 & 3 provide published studies indicating the role that parasite transmission can play in co-evolutionary models; when transmission is excluded important effects (dependent upon the interaction model and host ploidy) will be lost, potentially leading to false assumptions. Transmission, as a general process, can not be considered as equivalent when it is subject to different infection models. In the third chapter, strong interactions were present, and are observable (Figure 4.3), between transmission and the standard interaction models. These results should undermine the standard approach of the "single challenge" models, which ignore transmission in host-parasite co-evolutionary models. Along the lines of research by Lively (2010), but including the input of studies like that of Keeling (1999). A consequence of this on the RQH is that less support for its role on the evolution of recombination is forthcoming.

In chapter 1 investigation of the MTH and DMH lend credence to the non-adaptive evolution of male-haploidy. In particular, the role of the MTH (a form of genetic drive) was seen as the dominant mechanism explaining male-haploidy, under mild to severe costs, although this account of the strength of the MTH is biased from the use of a single locus viability model. Other factors

that might be involved in the evolution of male-haploidy could be the genetic system underlying sex-determination. In several cases, male-heterogametic systems would block the evolution of male-haploid (hemizygotic) genomes (Gardner & Ross, 2014). In the case of arrhenotoky the viable development of unfertilized eggs into fertile adults is likely to be crucial in its evolution.

Lastly, the use of stochastic models in chapters 1 & 3 highlight the role that variance can play in evolutionary models. For example, comparison of figures 1.2 & 1.4 from chapter 1 show that introducing stochasticity leads to less restrictive conditions (in selection and heterozygosity) for male-haploidy to invade. In chapter 3, the development of an agent based model allowed the incorporation of recombination as an individual level trait; the agents being individual hosts that are subject to the same population parameters (strength of selection, heterozygosity, viability costs and parasite transmission). In this scenario it is relatively simple to introduce a genetic model that allows for the recombination rate to evolve (it can be inherited and mutate between generations). This and similar situations are easy to incorporate in these types of model, potentially allowing for alternative avenues of research in evolutionary biology.

Bibliography

Agrawal, A. 2009. Differences between selection on sex versus recombination in red queen models with diploid hosts. *Evolution* **63**: 2131–2141.

Agrawal, A. & Lively, C. 2002. Infection genetics: gene-for-gene versus matching-alleles models and points in between. *Evolutionary Ecology Research* **4**: 79–80.

Agrawal, A. & Otto, S. 2006. Host-parasite coevolution and selection on sex through the effects of segregation. *The American Naturalist* **168**: 617–629.

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. *Proceedings of the Royal Society B* **278**: 3544–3550.

Asplen, M., Whitfield, J., Boer, J.D. & Heimpel, G. 2009. Ancestral state reconstruction analysis of hymenopteran sex determination mechanisms. *Journal of Evolutionary Biology* **22**: 1762–1769.

Baer, B. & Schmid-Hempel, P. 1999. Experimental variation in polyandry affects parasite loads and fitness in a bumble-bee. *Nature* **397**: 151–154.

Barton, N. 1995. A general model for the evolution of recombination. *Genetical Research* **65**: 123–144.

Beaumont, M., Zhang, W. & Balding, D. 2002. Approximate bayesian computation in population genetics. *Genetics* **162**: 2025–2035.

Beukeboom, L. 1995. Sex determination in hymenoptera - a need for genetic and molecular studies. *Bioessays* **17**: 813–817. doi:10.1002/bies.950170911.

Beukeboom, L., Kamping, A., Louter, M., Pijnacker, L., Katju, V., Ferree, P. & Werren, J. 2007. Haploid females in the parasitic wasp nasonia vitripennis. *Science* **315**: 206–206.

Beukeboom, L. & van de Zande, L. 2010. Genetics of sex determination in the haplodiploid wasp nasonia vitripennis (hymenoptera: Chalcidoidea). *Journal of genetics* **89**: 333–339. doi:10.1007/s12041-010-0045-7.

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

Borgia, G. 1980. Evolution of haplodiploidy: Models for inbred and outbred systems. *Theoretical Population Biology* **17**: 102–128.

Breeuwer, J. 1997. *Wolbachia* and cytoplasmic incompatibility in the spider mites *tetranychus urticae* and *t. turkestani*. *Heredity* **79**: 41–47.

Bull, J. 1979. An advantage for the evolution of male haploidy and systems with similar genetic transmission. *Heredity* **43**: 361–381.

Bull, J. 1985. Sex determining mechanisms: An evolutionary perspective. *Experientia* **41**: 1285–1296. doi:10.1007/BF01952071.

Campbell, B., Steffen-Campbell, J. & Gill, R. 1994. Evolutionary origin of whiteflies (hemiptera: Sternorrhyncha: Aleyrodidae) inferred from 18s rdna sequences. *Insect molecular biology* **3**: 73–88. doi:10.1111/j.1365-2583.1994.tb00154.x.

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.

Charlesworth, B. 1980. The cost of sex in relation to mating system. *Journal of Theoretical Biology* **84**: 655–671.

Charlesworth, B. & Charlesworth, D. 1998. Some evolutionary consequences of deleterious mutations. *Genetica* **102-103**: 3–19. doi:10.1023/A:1017066304739.

Craft, M. & Caillaud, D. 2011. Network models: An underutilized tool in wildlife epidemiology? *Interdisciplinary Perspectives on Infectious Diseases* pp. 1–12.

Crawley, M. 2005. *Statistics: An Introduction using R.* John Wily & Sons Ltd., West Sussex, UK.

Crow, J. & Kimura, M. 2009. *An Introduction to Population Genetics Theory.* The Blackburn Press, NJ.

Crozier, R. 1971. Heterozygosity and sex determination in haplo-diploidy. *American Naturalist* **105**: 399–&. doi:10.1086/282733.

de Boer, J., Groenen, M., Pannebakker, B., Beukeboom, L. & Kraus, R. 2015. Population-level consequences of complementary sex determination in a solitary parasitoid. *BMC Evolutionary Biology* **15**. doi:10.1186/s12862-015-0340-2.

de la Filia, A., Bain, S. & Ross, L. 2015. Haplodiploidy and the reproductive ecology of arthropods. *Current Opinion in Insect Science* **9**: 36–43. doi:http://dx.doi.org/10.1016/j.cois.2015.04.018.

Engelstädter, J. & Bonhoeffer, S. 2009. Red queen dynamics with non-standard fitness interactions. *PLOS computational biology* **5**: e1000469.

Engelstaedter, J. & Hurst, G. 2006. Can maternally transmitted endosymbionts facilitate the evolution of haplodiploidy? *Journal of evolutionary biology* **19**: 194–202.

Eyre-Walker, A. & Keightley, P. 1999. High genomic deleterious mutation rates in hominids. *Nature* **397**: 344–347.

Gardner, A. & Ross, L. 2014. Mating ecology explains patterns of genome elimination. *Ecology Letters* **17**: 1602–1612. doi:10.1111/ele.12383.

Gavrilov, I. 2007. A catalog of chromosome numbers and genetic systems of scale insects (homoptera: Coccinea) of the world. *Isr. J. Entomol* **37**: 1–45.

Gerber, H. & Klosterm, E. 1970. Sex control by bees. a voluntary act of egg fertilization during oviposition. *Science* **167**: 82 –&. doi:10.1126/science.167.3914.82.

Goday, C. & Esteban, M. 2001. Chromosome elimination in sciarid flies. *BioEssays* **23**: 242–250.

Goldstein, D. 1994. Deleterious mutations and the evolution of male haploidy. *The American Naturalist* **144**: 176–183.

Gullan, P. & Cook, L. 2007. Phylogeny and higher classification of the scale insects (hemiptera: Sternorrhyncha: Coccoidea). *Zootaxa* **1668**: 413–425.

Haig, D. 1993. The evolution of unusual chromosomal systems in sciarid flies: intragenomic conflict and the sex ratio. *Journal of Evolutionary Biology* **6**: 249–261.

Haigh, J. 1978. Accumulation of deleterious genes in a population - mullers ratchet. *Theoretical population biology* **14**: 251–267. doi:10.1016/0040-5809(78)90027-8.

Hamilton, M. 1967. Extraordinary sex ratios. *Scince* **156**: 477–488.

Hamilton, W. 1964. The genetical evolution of social behaviour. *Journal of Theoretical Biology* **7**: 1–16.

Haraguchi, Y. & Sasaki, A. 2000. The evolution of parasite virulence and transmission rate in a spatially structured population. *Journal of Theoretical Biology* **203**: 85–96.

Hartl, D. & Brown, S. 1970. The origin of male haploid genetic systems and their expected sex ratio. *Theoretical Population Biology* **1**: 165–190.

Hedrick, P., Philip, W. & Parker, J. 1997. Evolutionary genetics and genetic variation of haplodiploids and x-linked genes. *Annual Review of Ecology and Systematics* **28**: 55–83. doi: 10.1146/annurev.ecolsys.28.1.55.

Hedrick, P.W. 1994. Purging inbreeding depression and the probability of extinction: full-sib mating. *Heredity* **73**: 363–372.

Henter, H. 2003. Inbreeding depression and haplodiploidy: Experimental measures in a parasitoid and comparisons across diploid and haplodiploid insect taxa. *Evolution* **57**: 1793–1803.

Hill-Burns, E. & Clark, A. 2009. X-linked variation in immune response in *drosophila melanogaster*. *Genetics* **183**: 1477–1491.

Hodgeson, E. & Otto, S. 2012. The red queen coupled with directional selection favours the evolution of sex. *Journal of Evolutionary Biology* **25**: 797–802.

Howard, R. & Lively, C. 1994. Parasitism, mutation accumulation and the maintenance of sex. *Nature* **367**: 554–557.

Hughes, W. & Boomsma, J. 2006. Does genetic diversity hinder parasite evolution in social insect colonies. *Journal of Evolutionary Biology* **24**: 132–143.

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

Keen, N. 1990. Gene-for-gene complementarity in plant-pathogen interactions. *Annu. Rev. Genet.* **24**: 447–463.

Kidner, J. & Moritz, R. 2013. The red queen process does not select for high recombination rates in haplodiploid hosts. *Evolutionary Biology* **40**: 377–384.

Kidner, J. & Moritz, R. 2015. Host-parasite evolution in male-haploid hosts: an individual based network model. *Evolutionary Ecology* **29**: 93–105. doi:10.1007/s10682-014-9722-y.

King, K., Seppälä, O. & Neiman, M. 2012. Is more better? polyploidy and parasite resistance. *Biology Letters* **8**: 598–600.

Kuijper, B. & Pen, I. 2010. The evolution of haplodiploidy by male-killing endosymbionts: importance of population structure and endosymbiont mutualisms. *Journal of Evolutionary Biology* **23**: 40–52. doi:10.1111/j.1420-9101.2009.01854.x.

Liu, H., Zhang, X., Huang, J., Chen, J., Tian, D., Hurst, L. & Yang, S. 2015. Causes and consequences of crossing-over evidence via a high-resolution recombinational landscape of the honey bee. *Genome biology* **16**. doi:10.1186/s13059-014-0566-0.

Lively, C. 2010. An epidemiological model of host-parasite coevolution and sex. *Journal of Evolutionary Biology* **23**: 1490–1497.

Loewe, L. & Charlesworth, B. 2006. Inferring the distribution of mutational effects on fitness in drosophila. *Biology Letters* **2**: 426–430. doi:10.1098/rsb1.2006.0481.

Mable, B. & Otto, S. 1998. The evolution of life cycles with haploid and diploid phases. *BioEssays* **20**: 453–462.

Meznar, E., Gadau, J., Koeniger, N. & Rueppell, O. 2010. Comparative linkage mapping suggests a high recombination rate in all honeybees. *Journal of Heredity* **101**: s118–s126.

M'Gonigle, L. & Otto, S. 2011. Ploidy and the evolution of parasitism. *Proceedings of the Royal Society B* **278**: 2814–2822.

Misof, B., Liu, S., Meusemann, K., Peters, R., Donath, A., Mayer, C., Frandsen, P., Ware, J., Flouri, T., Beutel, R., Niehuis, O., Petersen, M., Izquierdo-Carrasco, F., Wappler, T., Rust, J., Aberer, A., Aspöck, U., Aspöck, H., Bartel, D., Blanke, A., Berger, S., Böhm, A., Buckley, T., Calcott, B., Chen, J., Friedrich, F., Fukui, M., Fujita, M., Greve, C., Grobe, P., Gu, S., Huang, Y., Jermiin, L., Kawahara, A., Krogmann, L., Kubiak, M., Lanfear, R., Letsch, H., Li, Y., Li, Z., Li, J., Lu, H., Machida, R., Mashimo, Y., Kapli, P., McKenna, D., Meng, G., Nakagaki, Y., Navarrete-Heredia, J., Ott, M., Ou, Y., Pass, G., Podsiadlowski, L., Pohl, H., von, B., Reumont, Schütte, K., Sekiya, K., Shimizu, S., Slipinski, A., Stamatakis, A., Song, W., Su, X., Szucsich, N., Tan, M., Tan, X., Tang, M., Tang, J., Timelthaler, G., Tomizuka, S., Trautwein, M., Tong, X., Uchifune, T., Walzl, M., Wiegmann, B., Wilbrandt, J., Wipfler, B., Wong, T., Wu, Q., Wu, G., Xie, Y., Yang, S., Yang, Q., Yeates, D., Yoshizawa, K., Zhang, Q., Zhang, R., Zhang, W., Zhang, Y., Zhao, J., Zhou, C., Zhou, L., Ziesmann, T., Zou, S., Li, Y., Xu, X., Zhang, Y., Yang, H., Wang, J., Wang, J., Kjer, K. & Zhou, X. 2014. Data from: Phylogenomics resolves the timing and pattern of insect evolution. *Science* doi:doi:10.5061/dryad.3c0f1.

Miyakawa, M. & A.S., M. 2015. Qtl mapping of sex determination loci supports an ancient pathway in ants and honey bees. *PLOS Genetics* **11**. doi:10.1371/journal.pgen.1005656.

Møller, A.P. 2006. A review of developmental instability, parasitism and disease: infection, genetics and evolution. *Infection, genetics and evolution* **6**: 133–140.

Moran, L., Schmidt, O., Gelarden, I., R.C. Parrish, I. & Lively, C. 2011. Running with the red queen: Host-parasite coevolution selects for biparental sex. *Science* **333**: 216–218. doi: 10.1126/science.1206360.

Nachman, M. & Crowell, S. 2000. Estimate of the mutation rate per nucleotide in humans. *Genetics* **156**: 297–304.

Niehuis, O., Gibson, J.D., Rosenberg, M.S., Pannebakker, B.A., Koevoets, T., Judson, A.K., Desjardins, C.A., Kennedy, K., Duggan, D., Beukeboom, L.W. et al. 2010. Recombination

and its impact on the genome of the haplodiploid parasitoid wasp nasonia. *PLoS One* **5**: e8597.

Normark, B. 2003. The evolution of alternative genetic systems in insects. *Annual Review of Entomology* **48**: 397–423.

Normark, B. 2004. Haplodiploidy as an outcome of coevolution between male-killing cytoplasmic elements and their hosts. *Evolution* **58**: 790–798.

Normark, B. 2006. Perspective: Maternal kin groups and the origins of asymmetric genetic systems—genomic imprinting, haplodiploidy, and parthenogenesis. *Evolution* **60**: 631–642.

Nuismer, S. & Otto, S. 2004. Host-parasite interactions and the evolution of ploidy. *PNAS* **101**: 11036–11039.

Oswald, B. & Nuismer, S. 2007. Neopolyploidy and pathogen resistance. *Proceedings of the Royal Society B* **274**: 2393–2397.

Ottersatter, M.C. & Thomson, J.D. 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. *Biological Journal of the Linnean Society* **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. *The American Naturalist* **154**: 393–405.

Peters, A. & Lively, C. 2007. Short- and long-term benefits and detriments to recombination under antagonistic coevolution. *Journal of Evolutionary Biology* **20**: 1206–1217.

Pijnacker, L., Ferwerda, M., Bolland, H. & Helle, W. 1980. Haploid female parthenogenesis in the false spider mite brevipalpus obovatus (acari: Tenuipalpidae). *Genetica* **51**: 211–214.

R Core Team 2013. *R: A Language and Environment for Statistical Computing*. R Fundation for Statistical Computing, Vienna, Austria.

Roode, J.D., Yates, A. & Altizer, S. 2008. Virulence-transmission trade-offs and population divergence in virulence in a naturally occuring butterfly parasite. *Proceedings of the National Academy of Science of the USA* **105**: 7489–7494.

Ross, L., Pen, I. & Shuker, D. 2010. Genomic conflict in scale insects: the causes and consequences of bizarre genetic systems. *Biological Reviews* **85**: 807–828. doi:10.1111/j.1469-185X.2010.00127.x.

Ross, L., Shuker, D., B.B., N. & Pen, I. 2012a. The role of endosymbionts in the evolution of haploid-male genetic systems in scale insects (coccoidea). *Ecology and Evolution* **2**: 1071–1081. doi:10.1002/ece3.222.

Ross, L., Shuker, D., Normark, B. & Pen, I. 2012b. The role of endosymbionts in the evolution of haploid-male genetic systems in scale insects (coccoidea). *Ecology and Evolution* **2**: 1071–1081. doi:10.1002/ece3.222.

Ross, L., Shuker, D. & Pen, I. 2011. The evolution and suppression of male suicide under paternal genome elimination. *Evolution* **65**: 554–563. doi:10.1111/j.1558-5646.2010.01148.x.

Salathe, M., Kouyos, R., Regoes, R. & Bonhoeffer, S. 2007. Rapid parasite adaptation drives selection for high recombination rates. *Evolution* **62**: 295–300.

Salathe, M., Kouyos, R.D. & Bonhoeffer, S. 2008. The state of affairs in the kingdom of the red queen. *Trends in Ecology & Evolution* **23**: 439–445. doi:10.1016/j.tree.2008.04.010.

Sasaki, K. & Obara, Y. 2002. Egg activation and timing of sperm acceptance by an egg in honeybees (apis mellifera l.). *Insectes Sociaux* **49**: 234–240. doi:10.1007/s00040-002-8307-x.

Schmid-Hempel, P. & Jokela, J. 2002. Socially structured populations and evolution of recombination under antagonistic coevolution. *The American Naturalist* **160**: 403–408.

Singh, N., Criscoe, D., Skolfield, S., Kohl, K., Keebaugh, E. & Schlenke, T. 2015. Fruit flies diversify their offspring in response to parasite infection. *Science* **349**: 747–750. doi:10.1126/science.aab1768.

Singh, N., Larracuente, A. & Clark, A. 2008. Contrasting the efficacy of selection on the x and autosomes in drosophila. *Molecular Biology and Evolution* **25**: 454–467.

Smith, N. 2000. The evolution of haplodiploidy under inbreeding. *Heredity* **84**: 186–192.

Solignac, M., Vautrin, D., Baudry, E., Mougel, F., Loiseau, A. & Cornuet, J. 2004. A microsatellite-based linkage map of the honeybee, *apis mellifera* l. *Genetics* **167**: 253–262.

Stahlhut, J. & Cowan, D. 2004. Single-locus complementary sex determination in the inbreeding wasp euodynerus foraminatus saussure (hymenoptera : Vespidae). *Heredity* **92**: 189–196. doi:10.1038/sj.hdy.6800394.

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 Genomics* **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. *Molecular Ecology* **20**: 4408–4420.

Szathmary, E. & Maynard, S. 1995. The major evolutionary transitions. *Nature* **374**: 227–232.

Takahashi, Y., Morita, S., Yoshimura, J. & Watanabe, M. 2011. A geographic cline induced by negative frequency-dependent selection. *BMC Evolutionary Biology* **11**: 256–266.

Tortajada, A., Carmona, M. & Serra, M. 2009. Does haplodiploidy purge inbreeding depression in rotifer populations? *PLoS ONE* **4**: 38195.

Úbeda, F. & Normark, B.B. 2006. Male killers and the origins of paternal genome elimination. *Theoretical Population Biology* **70**: 511 – 526. doi:http://dx.doi.org/10.1016/j.tpb.2006.06.011.

Ugelvig, L., Kronauer, D., Schrempf, A., Heinze, J. & Cremer, S. 2010. Rapid anti-pathogen response in ant societies relies on high genetic diversity. *Proceedings of the Royal Society B* **277**: 2821–2828.

van Baalen, M. & Beekman, M. 2006. The costs and benefits of genetic heterogeneity in resistance against parasites in social insects. *The American Naturalist* **167**: 568–577.

Vergara, D., Jokela, J. & Lively, C. 2014. Infection dynamics in coexisting sexual and asexual host populations: Support for the red queen hypothesis. *American Naturalist* **184**: s22–s30. doi:10.1086/676886.

Walker, T. & Hughes, W. 2009. Adaptive social immunity in leafcutting ants. *Biology Letters* **5**: 446–448.

Wallberg, A., Glemin, S. & Webster, M. 2015. Extreme recombination frequencies shape genome variation and evolution in the honeybee, apis mellifera. *PLOS genetics* **11**. doi:10.1371/journal.pgen.1005189.

Webb, S., Keeling, M. & Boots, M. 2013. A theoretical study of the role of spatial population structure in the evolution of parasite virulence. *Theoretical Population Biology* **84**: 36–45.

Werren, J. & Beukeboom, L. 1998. Sex determination, sex ratios and genetic conflict. *Annual review of Ecology and Systematics* **29**: 233–261.

West, S., Lively, C. & Read, A. 1999. A pluralist approach to sex and recombination. *Journal of evolutionary biology* **12**: 1003–1012.

Wilfert, L., Gadau, J. & Schmid-Hempel, P. 2007a. The genetic architecture of immune defense and reproduction in male bombus terrestris bumblebees. *Evolution* **61**: 804–815.

Wilfert, L., Gadau, J. & Schmid-Hempel, P. 2007b. Variation in genomic recombination rates among animal taxa and the case of social insects. *Heredity* **98**: 189–197.

Williams, T. & Kelly, C. 2003. *GNUPLOT: An Interactive Plotting Program.*

Zayed, A. 2004. Effective population size in hymenoptera with complementary sex determination. *Heredity* **93**: 627–630. doi:10.1038/sj.hdy.6800588.

Appendices

Appendix A

A.1 Frequency vectors

$$\mathbf{Q^f} = \left(\; D^2p^2, 2D^2pq, D^2q^2, 2Ddp^2, 4Ddpq, 2Ddq^2, d^2p^2, 2d^2pq, d^2q^2 \; \right) =$$

$$\mathbf{Q^f} = \left(\; X1, X2, X3, X4, X5, X6, X7, X8, X9 \; \right)$$

$$\mathbf{Q^m} = \left(\; D^2p^2, 2D^2pq, D^2q^2, (1-D^2)p, (1-D^2)q \; \right) =$$

$$\mathbf{Q^m} = \left(\; Y1, Y2, Y3, Y4, Y5 \; \right)$$

A.2 Selection vectors

$$\mathbf{S^f} = \left(\; 1, 1-hs, 1-s, 1, 1-hs, 1-s, 1, 1-hs, 1-s \; \right)$$

$$\mathbf{S^m} = \left(\; 1, 1-hs, 1-s, 1-c, (1-s)(1-c) \; \right)$$

A.3 The female inheritance matrices

The following matrices illustrate the proportion of genotype-genotype matings resulting in the relative genotype. Those values in bold **x** indicate those values that are used in the assortative mating matrices.

$$\mathbf{D^2p^2} = \left(\begin{array}{c|ccccccccc} & X1 & X2 & X3 & X4 & X5 & X6 & X7 & X8 & X9 \\ \hline Y1 & \mathbf{1.00} & 0.5 & 0 & \mathbf{0.5} & 0.25 & 0 & 0 & 0 & 0 \\ Y2 & 0.5 & \mathbf{0.25} & 0 & 0.25 & \mathbf{0.125} & 0 & 0 & 0 & 0 \\ Y3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ Y4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ Y5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

$$\mathbf{2D^2pq} = \left(\begin{array}{c|ccccccccc} & X1 & X2 & X3 & X4 & X5 & X6 & X7 & X8 & X9 \\ \hline Y1 & 0 & 0.5 & 1 & 0 & 0.25 & 0.5 & 0 & 0 & 0 \\ Y2 & 0.5 & \mathbf{0.5} & 0.25 & \mathbf{0.25} & 0.25 & 0 & 0 & 0 & 0 \\ Y3 & 1 & 0.5 & 0 & 0.5 & 0.25 & 0 & 0 & 0 & 0 \\ Y4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ Y5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

$$\mathbf{D^2q^2} = \left(\begin{array}{c|ccccccccc} & X1 & X2 & X3 & X4 & X5 & X6 & X7 & X8 & X9 \\ \hline Y1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ Y2 & 0 & \mathbf{0.25} & 0.5 & 0 & \mathbf{0.125} & 0.25 & 0 & 0 & 0 \\ Y3 & 0 & 0.5 & \mathbf{1} & 0 & 0.25 & \mathbf{0.5} & 0 & 0 & 0 \\ Y4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ Y5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

$$\mathbf{2Ddp^2} = \left(\begin{array}{c|ccccccccc} & X1 & X2 & X3 & X4 & X5 & X6 & X7 & X8 & X9 \\ \hline Y1 & 0 & 0 & 0 & \mathbf{0.5} & 0.25 & 0 & \mathbf{1} & 0.5 & 0 \\ Y2 & 0 & 0 & 0 & .25 & \mathbf{0.25} & 0 & 0.5 & \mathbf{0.5} & 0 \\ Y3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ Y4 & \mathbf{1} & 0.5 & 0 & \mathbf{0.5} & 0.25 & 0 & 0 & 0 & 0 \\ Y5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

$$\mathbf{4Ddpq} = \left(\begin{array}{c|ccccccccc} & X1 & X2 & X3 & X4 & X5 & X6 & X7 & X8 & X9 \\ \hline Y1 & 0 & 0 & 0 & 0 & 0.25 & 0.5 & 0 & 0.5 & 1 \\ Y2 & 0 & 0 & 0 & 0.25 & \mathbf{0.25} & 0.25 & 0.5 & \mathbf{0.5} & 0.5 \\ Y3 & 0 & 0 & 0 & 0.5 & 0.25 & 0 & 1 & 0.5 & 0 \\ Y4 & 0 & 0.5 & 1 & 0 & 0.25 & 0.25 & 0 & 0 & 0 \\ Y5 & 1 & 0.5 & 0 & 0.5 & 0.25 & 0 & 0 & 0 & 0 \end{array}\right)$$

$\mathbf{2Ddq^2} =$

	$X1$	$X2$	$X3$	$X4$	$X5$	$X6$	$X7$	$X8$	$X9$
$Y1$	0	0	0	0	0	0	0	0	0
$Y2$	0	0	0	0	**0.125**	0.25	0	**0.25**	0.5
$Y3$	0	0	0	0	0.25	**0.5**	0	0.5	**1**
$Y4$	0	0	0	0	0	0	0	0	0
$Y5$	0	0.5	**1**	0	0.25	**0.5**	0	0	0

$\mathbf{d^2p^2} =$

	$X1$	$X2$	$X3$	$X4$	$X5$	$X6$	$X7$	$X8$	$X9$
$Y1$	0	0	0	0	0	0	0	0	0
$Y2$	0	0	0	0	0	0	0	0	0
$Y3$	0	0	0	0	0	0	0	0	0
$Y4$	0	0	0	**0.5**	0.25	0	**1**	0.5	0
$Y5$	0	0	0	0	0	0	0	0	0

$\mathbf{2d^2pq} =$

	$X1$	$X2$	$X3$	$X4$	$X5$	$X6$	$X7$	$X8$	$X9$
$Y1$	0	0	0	0	0	0	0	0	0
$Y2$	0	0	0	0	0	0	0	0	0
$Y3$	0	0	0	0	0	0	0	0	0
$Y4$	0	0	0	0	0.25	0.5	0	0.5	1
$Y5$	0	0	0	0.5	0.25	0	1	0.5	0

$\mathbf{d^2q^2} =$

	$X1$	$X2$	$X3$	$X4$	$X5$	$X6$	$X7$	$X8$	$X9$
$Y1$	0	0	0	0	0	0	0	0	0
$Y2$	0	0	0	0	0	0	0	0	0
$Y3$	0	0	0	0	0	0	0	0	0
$Y4$	0	0	0	0	0	0	0	0	0
$Y5$	0	0	0	0	0.25	**0.5**	0	0.5	**1**

A.4 The male inheritance matrices

$\mathbf{D^2p^2} =$

	$X1$	$X2$	$X3$	$X4$	$X5$	$X6$	$X7$	$X8$	$X9$
$Y1$	**1**	0.5	0	**0.5**	0.25	0	0	0	0
$Y2$	0.5	**0.25**	0	0.25	**0.125**	0	0	0	0
$Y3$	0	0	0	0	0	0	0	0	0
$Y4$	0	0	0	0	0	0	0	0	0
$Y5$	0	0	0	0	0	0	0	0	0

$$\mathbf{2D^2pq} = \left(\begin{array}{c|ccccccccc} & X1 & X2 & X3 & X4 & X5 & X6 & X7 & X8 & X9 \\ \hline Y1 & 0 & 0.5 & 1 & 0 & 0.25 & 0.5 & 0 & 0 & 0 \\ Y2 & 0.5 & \mathbf{0.5} & 0.5 & 0.25 & \mathbf{0.25} & 0.25 & 0 & 0 & 0 \\ Y3 & 1 & 0.5 & 0 & 0 & 0.5 & 0.25 & 0 & 0 & 0 \\ Y4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ Y5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

$$\mathbf{D^2q^2} = \left(\begin{array}{c|ccccccccc} & X1 & X2 & X3 & X4 & X5 & X6 & X7 & X8 & X9 \\ \hline Y1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ Y2 & 0 & \mathbf{0.25} & 0.5 & 0 & \mathbf{0.125} & 0.25 & 0 & 0 & 0 \\ Y3 & 0 & 0.5 & \mathbf{1} & 0 & 0.25 & \mathbf{0.5} & 0 & 0 & 0 \\ Y4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ Y5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

$$\mathbf{(1 - D^2)p} = \left(\begin{array}{c|ccccccccc} & X1 & X2 & X3 & X4 & X5 & X6 & X7 & X8 & X9 \\ \hline Y1 & 0 & 0 & 0 & \mathbf{0.5} & 0.25 & 0 & \mathbf{1} & 0.5 & 0 \\ Y2 & 0 & 0 & 0 & 0.5 & \mathbf{0.25} & 0 & 1 & \mathbf{0.5} & 0 \\ Y3 & 0 & 0 & 0 & 0.5 & 0.25 & 0 & 1 & 0.5 & 0 \\ Y4 & 0 & 0 & 0 & \mathbf{0.5} & 0.25 & 0 & \mathbf{1} & 0.5 & 0 \\ Y5 & 0 & 0 & 0 & 0.5 & 0.25 & 0 & 1 & 0.5 & 0 \end{array}\right)$$

$$\mathbf{(1 - D^2)q} = \left(\begin{array}{c|ccccccccc} & X1 & X2 & X3 & X4 & X5 & X6 & X7 & X8 & X9 \\ \hline Y1 & 0 & 0 & 0 & 0 & 0.25 & 0.5 & 0 & 0.5 & 1 \\ Y2 & 0 & 0 & 0 & 0 & \mathbf{0.25} & 0.5 & 0 & \mathbf{0.5} & 1 \\ Y3 & 0 & 0 & 0 & 0 & 0.25 & \mathbf{0.5} & 0 & 0.5 & \mathbf{1} \\ Y4 & 0 & 0 & 0 & 0 & 0.25 & 0.5 & 0 & 0.5 & 1 \\ Y5 & 0 & 0 & 0 & 0 & 0.25 & \mathbf{0.5} & 0 & 0.5 & \mathbf{1} \end{array}\right)$$

A.5 Mutation matrices

$$\mathbf{M^m} = \left(\begin{array}{c|ccccc} & Y1 & Y2 & Y3 & Y4 & Y5 \\ \hline Y1 & (1 - 2\mu - \mu_d) & \nu & 0 & \nu_d & 0 \\ Y2 & 2\mu & (1 - \mu - \nu) & 2\nu & 0 & 0 \\ Y3 & 0 & \mu & (1 - 2\nu - \mu_d) & 0 & \nu_d \\ Y4 & \mu_d & 0 & 0 & (1 - \mu - \nu_d) & \nu \\ Y5 & 0 & 0 & \mu_d & \mu & (1 - \nu - \nu_d) \end{array}\right)$$

$$\mathbf{M^f} = \left(\begin{array}{c|ccc}
 & X1 & X2 & X3 \\ \hline
X1 & (1-2\mu-\mu_d) & \nu & 0 \\
X2 & 2\mu & (1-\mu-\nu-\mu_d) & 2\nu \\
X3 & 0 & \mu & (1-2\nu-\mu_d) \\
X4 & \mu_d & 0 & 0 \\
X5 & 0 & \mu_d & 0 \\
X6 & 0 & 0 & \mu_d \\
X7 & 0 & 0 & 0 \\
X8 & 0 & 0 & 0 \\
X9 & 0 & 0 & 0
\end{array}\right.$$

$$\left.\begin{array}{c|ccc}
 & X4 & X5 & X6 \\ \hline
X1 & \nu_d & 0 & 0 \\
X2 & 0 & \nu_d & 0 \\
X3 & 0 & 0 & \nu_d \\
X4 & (1-2\mu-\mu_d-\nu_d) & \nu & 0 \\
X5 & 2\mu & (1-\mu-\nu-\mu_d-\nu_d) & 2\nu \\
X6 & 0 & \mu & (1-2\nu-\mu_d-\nu_d) \\
X7 & \mu_d & 0 & 0 \\
X8 & 0 & \mu_d & 0 \\
X9 & 0 & 0 & \mu_d
\end{array}\right.$$

$$\left.\begin{array}{c|ccc}
 & X7 & X8 & X9 \\ \hline
X1 & 0 & 0 & 0 \\
X2 & 0 & 0 & 0 \\
X3 & 0 & 0 & 0 \\
X4 & \nu_d & 0 & 0 \\
X5 & 0 & \nu_d & 0 \\
X6 & 0 & 0 & \nu_d \\
X7 & (1-2\mu-\nu_d) & \nu & 0 \\
X8 & 2\mu & (1-\mu-\nu-\nu_d) & 2\nu \\
X9 & 0 & \mu & (1-2\nu-\nu_d)
\end{array}\right)$$

Appendix B

B.1 Pseudocode

This section contains the pseudocode for the core algorithms. We use the standard notation of leftwards pointing arrows indicating value assignment. When we use the leftwards pointing arrow after a sign such as +, or *times* the value contained by the left hand variable is incremented by the right hand variable. Another term for incrementation is the presence of ++ to the right, or left of a variable name, both have the effect of adding 1 to the value. The rightward pointing arrows indicate properties of the variable on the left side, this is generally used for frequencies. Lastly, when we use vectors or matrices we access the elements through the use of square brackets with 'array[0]' indicating the access of the first element of 'array'.

We first start by presenting algorithms that are not dependent. Here, calculating selection on the host genotypes is one of the lower level algorithms. The basic principle is to use the product of the infection probabilities (and mortality, inner most loop), to get the cumulative survival probabilities over all scenarios (second loop). The element $selection\ coefficient$ is the fraction indicating the number of matching loci between the host and parasite genotypes. While the $selection\ modifier$ element contains the parasite virulence parameter, multiplication with the $selectioncoefficient$ generates a range of virulence factors under an additive model. Under the two loci two allele model, the $selection\ coefficient$ coul take one of three values 1.0, 0.75 and 0.5.

The infection based process only incorporates the proportion of susceptible genotypes from the host population. Using the conditional frequencies allows for a reduction in the use of multiplications within the calculations. While a frequency based approach was attempted, conditions in the model tended to lead to the permanent extinction of the parasite strains. Hence a more proximate method was used with a much higher rate of transmission occurring (the *growth* variable) if the proportion of infected was much lower.

After selection and parasite growth, generation of the new host genotype frequencies was conducted. The first step in the process of this was calculating the frequencies of gamete genotypes from the particular parental frequencies. While calculation of the gamete frequencies was done per generation, the identification of the gamete proportions was done prior to execution

Algorithm 1 The algorithm for calculating selection on the host

for all host genotypes **do**
  $survival \leftarrow 0$
  for all host genotype $\rightarrow$ infection combinations **do**
    $selection \leftarrow preselection\ frequency$
    for all infectious parasite genotypes **do**
      if infected **then**
        $selection\times \leftarrow parasitefrequency \times (1 - selection\ modifier \times selection\ coefficient)$
      else
        $selection\times \leftarrow (1 - parasite\ frequency)$
      end if
    end for
    $survival+ \leftarrow selection$
  end for
end for

Algorithm 2 The algorithm for calculating the infection process

for all parasite genotypes **do**
  $infected \leftarrow 0$
  $uninfected \leftarrow 0$
  for all susceptible genotypes per parasite genotype **do**
    $infected+ \leftarrow hostfrequency \times parasitefrequency$
    $uninfected+ \leftarrow hostfrequency \times (1 - parasitefrequency)$ {We calculate the absolute probabilities from the conditional probabilities}
  end for
  $ratio = uninfected/infected$
  if $ratio \geq 2$ **then**
    $growth \leftarrow 4$
  else if $ratio$ ¡ 1 **then**
    $growth \leftarrow 2$
  else
    $growth\times \leftarrow 2$
  end if
end for

of the simulation. Hence in the following pseudocode one assumes correct identification of the recombinant fraction in addition to genotype. The end result of this algorithm is gamete frequencies per gender, while the gender is unimportant for the diploid scenario this is essential for the haplodiploid host.

Algorithm 3 The Method for Calculating Gametes

GametesProduced $\leftarrow 0$
GameteFrequencies $\leftarrow 0$
for $i = 0; i <$ no. adult genotypes contributing; $i++$ **do**
  if adult genotype[i]$\rightarrow$ homozygote **then**
    GametesProduced $\leftarrow 2$
  else
    GametesProduced $\leftarrow 1$
  end if
  if haploid **then**
    GameteFrequencies+ $\leftarrow$ adult genotype[i] $\rightarrow$ *PostSelectionFrequency* $\times$ *GametesProduced*
  else
    GameteFrequencies+ $\leftarrow$ adult genotype[i] $\rightarrow$ *PostSelectionFrequency* $\times$ *GametesProduced* $\times$ Stored Recombination Rate[i]
  end if
end for
save *GameteFrequencies* {This algorithm uses its own class, where each instance is designed for a particular sex. The gamete frequencies are temporarily stored for use later.}

After this the calculations simply follow filling in the 'punnet'

Algorithm 4 The recombination and gamete model

```
iterator = 0 {This iterator is used to go through each of the adult genotypes}
for i = 0; i < no. gamete genotypes; i++ do
  male gamete[i] → Calculate the Gamete Frequencies
  female gamete[i] → Calculate the Gamete Frequencies
end for
for i = 0; i < no. gamete genotypes; i++ do
  if male genotype[i] → Haploid then
    male genotype[i] → Frequency ← female gamete[i] → Frequency
  end if
  for j = 0; j < no. gamete genotypes; j++ do
    if male genotype[iterator] → Diploid then
      if i ≠ j then
        male genotype[iterator] → Frequency ← male gamete[i] → Frequency ×
        female gamete[j] → Frequency + female gamete[i] → Frequency × male gamete[j] →
        Frequency
      else
        male genotype[iterator] → Frequency ← male gamete[i] → Frequency ×
        female gamete[i] → Frequency
      end if
    end if
    if i ≠ j then
      female genotype[iterator] → Frequency ← male gamete[i] → Frequency ×
      female gamete[j] → Frequency + female gamete[i] → Frequency × male gamete[j] →
      Frequency
    else
      female genotype[iterator] → Frequency ← male gamete[i] → Frequency ×
      female gamete[i] → Frequency
    end if
    iterator++ {Increment the iterator by a value of 1}
  end for
end for
```

B.2 Algebraic analysis of infection dynamics

We first start by defining the change in frequency of the infected proportion of the susceptible population. We do so with a simplified model system (as the sustainability of the system is not a concern here).

$$\Delta f\left(I\right) = \frac{f\left(I_0\right) R_0 - f\left(I_0\right)}{f\left(S_0\right)} \tag{B.1}$$

Where we define $f\left(I_0\right)$ as the proportion, out of the total population, infected by the parasite strain at time 0. R_0 is used to indicate the growth rate of the parasite strain (for the parasite strain to spread this value needs to be greater than 1), as per standard infection models. Lastly we need to define the proportion of the population that is susceptible to the parasite strain, we use $f\left(S_0\right)$ to indicate this frequency at time 0. To get the change in frequency we need to divide the difference between the proportion of the host population infected before and after the infection process, by the frequency of susceptible individuals in the population. Upon doing so we get the conditional change in probability.

To get any further we need to define the term R_0: Here it is defined as.

$$R_0 = \frac{f(I_0) + f(I_0) \cdot (1 - f(I_0)) \cdot f(S_0)}{f(I_0)} \tag{B.2a}$$

$$R_0 = (1 + (1 - f(I_0)) \cdot f(S_0)) \cdot f(S_0) \tag{B.2b}$$

We use the derived equation 2a to be integrated with equation 1. The idea in the above two equations (more obvious in 2a) is the probability of an infected individual (which occurs with probability $f(I_0)$) meeting an uninfected susceptible individual $(1-f(I_0))\cdot f(S_0)$. While we could alter the density of the interactions the main point is the relation of the values of $f(S_0)$ with $f(I_0)$, due to this we leave the function as simple as this. However, as we use the denominator of $f(S_0)$ in equation 1 to get a conditional probability, the factor R_0 (in equation 1) needs to be an absolute value, hence we need to multiply equation 2a by $f(S_0)$. Subsequently we integrate equation 2b into equation 1.

$$\Delta f(I) = f(I_0)(1 + (1 - f(I_0)) \cdot f(S_0)) \cdot f(S_0) - f(I_0) \tag{B.3a}$$

$$= f(I_0) \cdot (f(S_0) - 1 + f(S_0)^2(1 - f(I_0))) \tag{B.3b}$$

$$f(S_0)(1 + f(S_0)) - (1 + f(S_0)^2 f(I_0)) \tag{B.3c}$$

Equation 3a is equation 1 with equation 2b integrated. After expansion and rearrangement we eventually arrive at equation 3b. At this point we ignore the term $f(I_0)$ that is outside of the main block, all we are really interested in is whether the parasite spreads or not. Doing this then leaves us with an equality, of which one (before being established) term can be considered to be

more or less insignificant ($f(S_0)^2 \cdot f(I_0)$). At this point we can either run through permutations of genotype frequencies, to establish when parasites can invade a host population. The result is that for a parasite to successfully invade a population $\approx 61\%$ of the population needs to be susceptible. In this situation the susceptible haplotype frequency in the population is 0.3053 in the diploid host population and 0.4524 in the male-haploid population.

B.3 Infection matrices

Table B.1: The infection matrix for host genotype AB/AB

	uninfected(0)	(1)
AB	$1 - p(P_r)$	$p(P_r) \cdot (1 - s_r)$

Table B.2: The infection matrix for host genotype aB/aB

	uninfected(0)	(1)
aB	$1 - p(P_r)$	$p(P_r) \cdot (1 - s_r)$

Table B.3: The infection matrix for host genotype Ab/Ab

	uninfected(0)	(1)
Ab	$1 - p(P_r)$	$p(P_r) \cdot (1 - s_r)$

Table B.4: The infection matrix for host genotype ab/ab

	uninfected(0)	(1)
ab	$1 - p(P_r)$	$p(P_r) \cdot (1 - s_r)$

Table B.5: The infection matrix for host genotype AB/aB

	uninfected(0,0)	(1,0)	(0,1)	(1,1)
AB	$1-p(P_r)$	$p(P_r)\cdot(1-s_r)$	$1-p(P_r)$	$p(P_r)\cdot(1-s_r)$
aB	$1-p(P_r)$	$1-p(P_r)$	$p(P_r)\cdot(1-s_r)$	$p(P_r)\cdot(1-s_r)$

Table B.6: The infection matrix for host genotype AB/Ab

	uninfected(0,0)	(1,0)	(0,1)	(1,1)
AB	$1-p(P_r)$	$p(P_r)\cdot(1-s_r)$	$1-p(P_r)$	$p(P_r)\cdot(1-s_r)$
Ab	$1-p(P_r)$	$1-p(P_r)$	$p(P_r)\cdot(1-s_r)$	$p(P_r)\cdot(1-s_r)$

Table B.7: The infection matrix for host genotype ab/aB

	uninfected(0,0)	(1,0)	(0,1)	(1,1)
ab	$1-p(P_r)$	$p(P_r)\cdot(1-s_r)$	$1-p(P_r)$	$p(P_r)\cdot(1-s_r)$
aB	$1-p(P_r)$	$1-p(P_r)$	$p(P_r)\cdot(1-s_r)$	$p(P_r)\cdot(1-s_r)$

Table B.8: The infection matrix for host genotype ab/Ab

	uninfected(0,0)	(1,0)	(0,1)	(1,1)
ab	$1-p(P_r)$	$p(P_r)\cdot(1-s_r)$	$1-p(P_r)$	$p(P_r)\cdot(1-s_r)$
Ab	$1-p(P_r)$	$1-p(P_r)$	$p(P_r)\cdot(1-s_r)$	$p(P_r)\cdot(1-s_r)$

Table B.9: The infection matrix for host genotype AB/ab

	uninfected(0,0,0,0)	(1,0,0,0)	(0,1,0,0)	(0,0,1,0)
AB	$1-p(P_r)$	$p(P_r)\cdot(1-s_r)$	$1-p(P_r)$	$1-p(P_r)$
Ab	$1-p(P_r)$	$1-p(P_r)$	$p(P_r)\cdot(1-s_r)$	$1-p(P_r)$
aB	$1-p(P_r)$	$1-p(P_r)$	$1-p(P_r)$	$p(P_r)\cdot(1-s_r)$
ab	$1-p(P_r)$	$1-p(P_r)$	$1-p(P_r)$	$1-p(P_r)$
	(0,0,0,1)	(1,1,0,0)	(1,0,1,0)	(1,0,0,1)
AB	$1-p(P_r)$	$p(P_r)\cdot(1-s_r)$	$p(P_r)\cdot(1-s_r)$	$p(P_r)\cdot(1-s_r)$
Ab	$1-p(P_r)$	$p(P_r)\cdot(1-s_r)$	$1-p(P_r)$	$1-p(P_r)$
aB	$1-p(P_r)$	$1-p(P_r)$	$p(P_r)\cdot(1-s_r)$	$1-p(P_r)$
ab	$p(P_r)\cdot(1-s_r)$	$1-p(P_r)$	$1-p(P_r)$	$p(P_r)\cdot(1-s_r)$
	(0,1,1,0)	(0,1,0,1)	(0,0,1,1)	(1,1,1,0)
AB	$1-p(P_r)$	$1-p(P_r)$	$1-p(P_r)$	$p(P_r)\cdot(1-s_r)$
Ab	$p(P_r)\cdot(1-s_r)$	$p(P_r)\cdot(1-s_r)$	$1-p(P_r)$	$p(P_r)\cdot(1-s_r)$
aB	$p(P_r)\cdot(1-s_r)$	$1-p(P_r)$	$p(P_r)\cdot(1-s_r)$	$p(P_r)\cdot(1-s_r)$
ab	$1-p(P_r)$	$p(P_r)\cdot(1-s_r)$	$p(P_r)\cdot(1-s_r)$	$1-p(P_r)$
	(1,1,0,1)	(1,0,1,1)	(0,1,1,1)	(1,1,1,1)
AB	$p(P_r)\cdot(1-s_r)$	$p(P_r)\cdot(1-s_r)$	$1-p(P_r)$	$p(P_r)\cdot(1-s_r)$
Ab	$p(P_r)\cdot(1-s_r)$	$1-p(P_r)$	$p(P_r)\cdot(1-s_r)$	$p(P_r)\cdot(1-s_r)$
aB	$1-p(P_r)$	$p(P_r)\cdot(1-s_r)$	$p(P_r)\cdot(1-s_r)$	$p(P_r)\cdot(1-s_r)$
ab	$p(P_r)\cdot(1-s_r)$	$p(P_r)\cdot(1-s_r)$	$p(P_r)\cdot(1-s_r)$	$p(P_r)\cdot(1-s_r)$

Appendix C

C.1 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 insectsOttersatter & Thomson (2007)), 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 5 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 6 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

Parasite transmission (Algorithms 4–6)

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 Keeling (1999).

Algorithm 7 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($current\_node, input\_variables$);
$no\_edges \leftarrow min\_edge\_no + \text{Random}(max\_edge\_no - min\_edge\_no)$;
Add edges(no_edges);

Algorithm 8 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**
  $infect \leftarrow no\_susceptible$;
  $Statistics \rightarrow \text{Update}(no\_susceptible)$;
else if $infect = 0$ **then**
  $infect \leftarrow 1$;
  $Statistics \rightarrow \text{Update}(1)$;
end if
$i \leftarrow 0$;
while $i < infect$ **do**
  $hosts_{iterators_i} \rightarrow \text{Infect(this)}$;
  $i \leftarrow i + 1$;
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

Algorithm 9 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
```

Algorithm 10 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 ← 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
```

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 9–14)

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

Algorithm 11 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 12 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

two locus system employed here.

Algorithm 13 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 $\leftarrow$ Random(2.0);
possibilities $\leftarrow 0$;
recombination the recombination rate of the current entity {Holds the cumulative value of the probabilities}
if *no_Genotypes* > 1 **then**
  for all *Genotypes* **do**
    $Recombinant\_events \leftarrow Genotype \rightarrow$ 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 14 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}
$recombination \leftarrow (paternal \rightarrow recombination + maternal \rightarrow recombination)/2 +$ Mutate();
{*recombination* is kept within the bounds; 0, 0.5. Mutate() generates variance around a mean of 0}
for all $maternal \rightarrow loci$ **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 15)

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 15 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

C.2 Interaction matrices

Table C.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
$2ab/aB$	1	1	0	0
$2ab/Ab$	1	0	1	0
$2ab/AB$	1	1	1	1
aB/aB	0	1	0	0
$2aB/Ab$	1	1	1	1
$2aB/AB$	0	1	0	1
Ab/Ab	0	0	1	0
$2Ab/AB$	0	0	1	1
AB/AB	0	0	0	1

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

C.3 Table C.2

Table C.2: The number of infected individuals observed at the peak of a parasite outbreak (the results are taken from a single simulation run). The results shown are the median values (and the range of observed values). The number of infected individuals is much higher under the MAM (a situation that favours haploid parasites unlike the IMAM). Comparisons between the different host population types shows consistently higher infection rates under conditions favouring higher infected - susceptible ratios.

MAM	IMAM	
Diploid	450 (353 – 699)	124.5 (105 – 144)
Haplodiploid	409.5 (206 – 696)	169.5 (109 – 280)