Evolutionary Genetics of Adaptation in Lepidoptera

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor of Philosophy

By Louise Reynolds

May 2016

Cover photo credit: L. Reynolds

Abstract

Revolutions in sequencing technology have provided an unprecedented opportunity to uncover the genetic basis of traits of adaptive importance, enabling researchers not only to merely describe the means of inheritance of traits but also to establish the genetic changes under selection. This thesis examines the loci involved in two recent episodes of strong selection, namely the suppression of *Wolbachia*-induced male-killing in the butterfly *Hypolimnas bolina* and the evolution of melanism in moths. *H. bolina* has evolved the ability to suppress the male-killing effects of the heritable endosymbiont *Wolbachia.* This thesis demonstrates firstly that this results from a single locus trait that doesn't involve genetic variants beyond chromosome 25. This simple genetic basis explains in part the speed of spread of the suppressor in natural populations. The hypothesis that the insect sex determination gene *doublesex* is the target of selection is then examined. Compatible with this hypothesis, *doublesex* variants were found to cosegregate with suppression, and that the peak of a selective sweep is located within *doublesex* region. An unusual pattern of inheritance was uncovered at the *doublesex* locus, suggestive of a duplication event. These data are consistent with, though not proof of, *Wolbachia* driving the evolution of this key sex determination gene. The progression of a selective sweep for the suppression, as it travelled in space across Independent Samoa, was then examined. The sweep across Independent Samoa corroborated the genomic region immediately around *doublesex* as the target of selection. The sweep was very broad but weakened as it progressed across Samoa. The thesis then examines the genetic basis of melanism in Lepidoptera, and compares the genomic region associated with a naturally selected melanistic form to a laboratory mutant. The 'natural' example corresponded to a known genomic hotspot for colour pattern evolution, whereas this region was excluded in the laboratory mutants. These data support the pleiotropic view of convergence – that involvement of a single region is associated with minimized non-target effects. The thesis ends with a discussion of these data and a programme for future research in the area.

Contents

Abbreviations

bp	Base pair
BLAST	Basic local alignment search tool
cM	CentiMorgan
CMS	Cytoplasmic male sterility
dsx	*doublesex*
RAD-seq	Restriction site Associated DNA sequencing
EtOH	Ethanol
Indel	Insertion/deletion
LD	Linkage disequilibrium
MC1R	Melanocortin 1 receptor
NGS	Next generation sequencing
ORF	Open reading frame
PCR	Polymerase chain reaction
rpm	Revolutions per minute
SAP	Shrimp alkaline phosphatase
SNP	Single nucleotide polymorphism
Ta	Annealing temperature

Chapter 1

Introduction

The origins and diversity of life have long been of interest to man. Charles Darwin and Alfred Russell Wallace revolutionised our understanding of the origins of this diversity with their theory of natural selection (Darwin & Wallace 1858). Ever since this seminal contribution, the goal of evolutionary biologists has been to understand the evolutionary processes that generate "endless forms most beautiful" (Darwin 1859).

Darwinian theory was then synthesized with Mendelian genetics to create the 'New Synthesis'. Within this body of theory, the dynamics of genetic variants with known mechanisms of inheritance, dominance and interactions with other genes was tracked to explicitly model how the genetic variants would spread (and recede) through the process of natural selection (Fisher 1930). To these selective processes were added stochastic elements based on sampling and genetic drift processes (Wright 1931; Wright 1932). Finally, where genetics could not be explicitly incorporated, quantitative genetic analysis was developed as a means of modelling evolutionary processes, with trait variation simply partitioned into variance associated with additive and epistatic components, and the interaction of these with the environment (Falconer 1952).

This body of work produces an excellent description of the behaviour of genetic variation over evolutionary time. However, there is no requirement within this framework to understand the actual mutations and genes involved in the evolutionary process. Largely as a matter of necessity, most research into phenotypic diversity and its origins has been concerned with determining whether a given trait is adaptive and the reasons for this adaptation, without ever pinpointing the molecular diversity responsible for the trait. Understanding of the genetic basis of diversity has predominately been developed within the field of population genetics, which has been able to predict and explain how variation is established and maintained without necessarily identifying the exact molecular mechanisms that underpin and control phenotypes.

The advent and maturation of Next Generation Sequencing (NGS) techniques has radically transformed our ability to seek out and identify the loci responsible for phenotypic variation. It is becoming increasingly possible to locate the loci of evolutionary change and

adaptation, with an ever growing database of adaptive traits for which the loci responsible have been identified (Martin & Orgogozo 2013; Pardo-Diaz et al. 2015). Beyond this, the advent of CRISPR-CAS technology allows defined testing of the functional role of particular mutations within the loci identified. Commonly used already in *Drosophila* and other model organisms (Bassett et al. 2013; Gratz et al. 2013; Yu et al. 2013), this technique is expanding to be applicable to all laboratory model organisms. Thus, the field of genomic analysis, both in terms of sequencing and functional studies, gives a previously unobtainable ability to establish the genetic basis of traits under selection in terms of both loci, and latterly, particular mutational events.

Now that the technology to locate the loci of adaptive traits has become more accessible, there is a temptation to define the loci and mutation for all traits of interest. This has fuelled a debate as to whether this process is worthwhile. Is it the case that identifying the molecular basis of phenotypic variation and striving towards an ever more complete description of the underlying mutations always enriches our understanding of evolutionary processes? Alternatively, are such endeavours merely the completion of evolutionary stories to molecular detail, but which actually contribute little of scientific value in terms of understanding the processes that give rise to diversity? (Rausher & Delph 2015; Travisano & Shaw 2013). To summarise the latter view, describing the loci and mutations underlying evolution is only worthwhile when it actually enriches our view of evolutionary processes.

In this thesis, I present two case studies examining the molecular basis of evolutionary changes. The first set of studies examines a particularly strong episode of selection observed recently in natural populations, namely selection for suppression of the male-killing activity of *Wolbachia* in a tropical nymphalid butterfly, *Hypolimnas bolina*. The justification for 'molecular discovery' in this case is twofold. First, the intense nature of selection may drive changes in systems that would otherwise remain static, as changes would ordinarily be deleterious. Second, the nature of the selective factor – death of males – produces a direct hypothesis as to the target of selection – a component of the sex determination system (Cordaux et al. 2011). The second study system being the somewhat more traditional system of the evolution of melanism in Lepidoptera. The spread of dark wing pattern forms in Lepidoptera associated with Industrial pollution and soot deposition has been observed in multiple moth species (Majerus 1998). This parallel evolution of a single phenotypic trait

produces the opportunity to investigate whether parallel evolutionary events in different species are reflected in convergence in the genetic basis of the trait. Further, if there is convergence, why does this exist in terms of the gene networks underlying phenotype?

Below, the conceptual basis underlying these questions is outlined in more detail, alongside a brief history of the study systems used in order to address these questions.

Question 1: Sex ratio distorting symbionts and the evolution of sex determination systems

Heritable symbionts – microbes that pass matrilineally from a female to her progeny – are common in insects, with current estimates indicating that the majority of insect species carry one or more heritable microbes (Zug & Hammerstein 2012). The interactions between heritable symbionts and their hosts are diverse both in terms of species identity, and the impact that the symbiont has upon the host.

a) Identity of symbiont: Heritable symbionts can be viral, bacterial, or single cell eukaryotes such as members of the Microspora. Within bacteria, there are multiple evolutionary transitions to becoming a heritable microbe within an arthropod, spanning multiple bacterial divisions. Whilst *Wolbachia* represents the best known symbiont (by virtue of its commonness in arthropods and its diverse effects), heritable *Rickettsia*, *Spiroplasma*, *Cardinium* and *Arsenophonus* symbionts are all found in >5% of arthropod species (Duron et al. 2008). To these can be added 'sporadic' heritable symbiont infections that are found narrowly in one group of insects – e.g. the gammaproteobacteria *Buchnera*, *Hamiltonella*, *Regiella* in aphids (Moran et al. 2008).

b) Interaction with the host: The impact that a symbiont has on its host is ascertained through a combination of association of infection with phenotypic variation (where uninfected and infected members of a population coexist), examination of the impact of antibiotic elimination, and genomic and physiological analyses. Such analyses have revealed there to be multiple types of interaction. In some cases, heritable microbe presence is required – antibiotic treatment causes host sterility or inviability. Here, the symbiont is either contributing to host function through provision of essential nutrients e.g. *Buchnera* in

aphids, or the host has evolved to be dependent on symbiont presence, whose elimination then leads to failure of host systems e.g. (Dedeine et al. 2001).

In other cases the elimination of a heritable microbe leaves a viable, fertile host. In such instances, the symbionts have been found to provide more subtle modulations of host biology that improve host fitness, or to act as reproductive parasites, or on occasions, have both phenotypes. In terms of beneficial effects, heritable microbes may for instance represent an important means of defence against natural enemies (Haine 2008). Alternatively, heritable microbes may display reproductive parasitic phenotypes. Heritable microbes are commonly transmitted matrilineally through the female line alone. This associates microbe fitness with that of their female, but not male, host. This association has driven the evolution of a variety of manipulations of host biology (reproductive parasitisms) that increase the production or survival of female hosts. This is evidenced in the presence of heritable microbes that induce parthenogenesis, feminize and kill males, in addition to strains that induce the phenotype of cytoplasmic incompatibility (CI), see (Hurst & Frost 2015) for review.

Reproductive parasitism: a brief introduction. Whilst the means through which reproductive parasites spread through populations has been understood for many years, the mechanisms by which these symbionts produce their parasitic phenotype are less well known. The microbes involved in reproductive parasitism are refractory to culture in the lab and thus not amenable to functional genetic analysis. The host systems upon which these parasites act are better studied. For instance, it is known that feminization of *Armadillidium* by *Wolbachia* functions through prevention of formation of the androgenic gland. Parthenogensis induction works through diverse mechanisms, including restoration of diploidy in arrhenotokous species (Hurst et al. 1999).

Male-killing works in diverse ways. In *Drosophila nebulosa, Spiroplasma*-induced male-killing is known to be associated with wide-spread apoptosis (Bentley et al. 2007). This process requires a fully-functioning dosage compensation complex in order for *Spiroplasma poulsonii* to effectively induce male host death (Veneti et al. 2005). In individuals with a male-like dosage compensation complex, *Spiroplasma* disrupts neuronal and cellular organisation during embryogenesis, possibly through the secretion of toxins (Ventura et al.

2012). *Arsenophonus nasoniae* causes male-killing in *Nasonia vitripennis* by preventing infected embryos from forming maternally-derived centrosomes, which are crucial for male, but not female, development (Ferree et al. 2008). In *Drosophila bifasciata*, male-killing has been shown to be caused by *Wolbachia*-induced interference of chromatin remodelling in males (Riparbelli et al. 2012). *Wolbachia* male-killing in *Ostrinia* is associated with female-like splicing of the gene *doublesex* in karyotypically male individuals (Sugimoto *et al*. 2012).

Reproductive parasitism causes strong fitness loss to the host. The deviation of population sex ratio away from a ratio of 1 female per male that is associated with the spread of a feminizing or male-killing element creates Fisherian selection for restoration of the sex ratio. For male-killing, there is an additional selective cost associated with the death of males. These effects combine to favour mutations in the host that rescue the wild type phenotype. These may be mutations that prevent heritable microbe transmission e.g. (Rigaud & Juchault 1992) or prevent heritable microbe action. Evolution of the latter are most commonly observed and are termed suppressors – mutations that prevent the microbial manipulation phenotype. Suppressors of male-killing have been directly observed in the butterfly *Hypolimnas bolina* and the ladybird *Menochilus sexmaculatus* (Majerus & Majerus 2010a). In other cases, suppressor presence has been inferred through emergence of male-killing phenotype following transinfection (Fujii et al. 2001; Sasaki et al. 2002).

The genetic nature of these suppressors of male-killing have not yet been established in either case. They are of interest because selection for suppression can be very strong – allowing mutations to spread that are otherwise deleterious. Thus, selective events associated with suppression may allow a species to cross an otherwise deep adaptive valley. Further, the targeting of male-killers of host sex determination systems creates the hypothesis that it is sex determination systems that may evolve means of suppression. Reproductive parasite involvement in sex determination systems has already been established as a driver of many transitions to asexuality, and in changes from karyotypic to symbiont-derived sex determination, see (Cordaux et al. 2011) for review. In the suppressor hypothesis, a further cause of evolution of sex determination systems is alteration in the system to prevent male-killer action.

Study system: Hypolimnas bolina – Wolbachia interaction

Hypolimnas bolina (Linnaeus, 1758), commonly known as the Great/Common/Varied Eggfly or Blue Moon Butterfly, is a member of the family Nymphalidae. A tropical butterfly, *H. bolina* is commonly found throughout South-East Asia and the South Pacific, with an endemic latitudinal distribution that ranges from Madagascar in the West (Clarke et al. 1975) to the Pitcairn islands in the East (Tennent 2006) and a longitudinal distribution that ranges from southern China in the north to Australia in the south (Clarke & Sheppard 1975). *H. bolina* is however highly migratory, often being found as far north as Japan (Mitsuhashi et al. 2011) and as far south as New Zealand (Patrick 2004; Ramsay 1971; Ramsay & Ordish 1966; Rosenberg 1973; Winstanley 1982). Furthermore, Clarke & Sheppard (1975) reported that it has been found as far East as Rapa Nui, although more recent reports have not recorded any such instance on this island (Dunn 2007).

H. bolina is sexually dimorphic. Whereas a single male morph is found throughout the entire species range, the female is polymorphic, with four basic morphs (Clarke & Sheppard 1975; Poulton 1924). Forms *naresi, pallescens* and *nerina* are found to the east of the *H. Bolina* distribution whilst *f. Euploeoides* is found to the west of the range (Clarke & Sheppard 1975). The euploeoides morph is controlled by a different locus to the other *H. Bolina* morphs, it forms part of the Euploea mimicry ring, alongside *E. Core*, *E. Sylvester*, *E. Klugii* and the Papilionidae *P. Clytia*, *f. Clytia* and *P. Dravidarum*, with female *H. Bolina* f.*euploeoides* closely resembling the local Euploea species (Clarke & Sheppard 1975).

The discovery that females in some parts of the *H. bolina* range produce all-female broods was deemed “so surprising” by the established entomologist Poulton “that it was desirable to record the fact without delay” and so he promptly presented Simmonds’ “astonishing results with this remarkable butterfly” (Poulton, 1923 p.ix) to the ‘Entomological Society of London’ even before a detailed account could be published (Poulton 1923; Simmonds 1926). Further investigations by Clarke and Sheppard recapitulated Simmonds’ earlier finding that this “all-female trait” was maternally-inherited, could persist for many generations, and that the 50% egg hatch rate in these female-biased lines did not result from parthenogenesis (Clarke et al. 1975; Poulton 1923; Poulton 1928; Simmonds 1926; Simmonds 1931). Clarke &

Sheppard (1975) also went on to establish that the lack of male offspring resulted from male-death occurring during the embryogenic and early larval phases. They hypothesised that all-female broods resulted from a cytoplasmic factor, which they were unable to identify (Clarke et al. 1975). This cytoplasmic factor was later identified as a male-killing strain of *Wolbachia* wBol1 (Dyson et al. 2002). At the individual level male-killing causes all-female broods (Clarke et al. 1975; Dyson et al. 2002; Poulton 1923), which can in turn lead to female-biased sex ratios at the population level (Dyson & Hurst 2004). These sex ratios can be extreme (Hamilton 1967). Such is the case for *H. bolina* populations on Independent Samoa. Here, a sex ratio bias so extreme that it led to a change in mating behaviour (Charlat et al. 2007c) persisted for more than a century (Dyson & Hurst 2004).

As selection for sex ratio is frequency-dependent (Fisher 1930) any mechanism that moves the population sex ratio away from parity will cause selection to act in favour of the rarer sex. The first indications that suppressors of male-killing existed in *H. bolina* were the presence of *w*Bol1-infected males in South-East Asia (Charlat et al. 2006; Charlat et al. 2005). In these populations, crossing data indicated *H. bolina* possesses a mutation that suppresses the male-killing effect of *Wolbachia* infection (Charlat et al. 2006; Hornett et al. 2009).

Should a suppressor evolve, which rescues males, then it would be strongly selected for and could cause rapid fluxes in sex ratio (Charlat et al. 2007a; Hamilton 1967; Hornett et al. 2009; Randerson et al. 2000). This is the case for the *H. bolina* populations of Independent Samoa. Here, a female-biased sex ratio of 100 females for every male had persisted for some one hundred years. Between 2001 and 2005 however, the sex ratio of this population had transitioned to a 1:1 sex ratio on the island of Upolu, one of the two main islands that constitutes Independent Samoa. The transition was then seen to occur on the neighbouring island of Savai'i. Breeding analysis revealed that *w*Bol1-infected females sampled after 2006 produced sex ratios near to 1:1 without male death. Furthermore, this change in sex ratio was shown to result from host genetic changes, as inferred by the reappearance of male-killing when the Independent Samoan wBol1 infection was bred onto a genetic background from a population in which male-killing was still present (Charlat et al. 2007a).

Later analysis of the genetic basis of the suppression trait localised it to a 10cM region of chromosome 25 (*Bombyx mori* equivalent). However, many features of the genetic basis of suppression were left unresolved. First, the involvement of other chromosomes in suppression was not excluded. Second, the region defined through mapping as carrying the suppressor was noted to include the gene *doublesex* in other Lepidoptera. *doublesex* is a nexus gene within the insect sex determination system, defining male and female identity at the cellular level through differential splicing. *doublesex* splicing is modulated by male-killing *Wolbachia* in *Ostrinia* moths (Sugimoto & Ishikawa 2012), making it an excellent candidate system for involvement in suppression of male-killing. However, the position of *doublesex* in the *H. bolina* genome and its association with suppression of male-killing were uncertain.

Question 2: Melanism in Lepidoptera as a convergent trait

Convergent evolution, whereby different species independently evolve similar phenotypes under similar selection pressures (Arendt & Reznick 2008) can be observed throughout nature. Observation of such phenomenon at a phenotypic level has given rise to the question of whether convergence at the level of the phenotype is also accompanied by convergence at the level of the molecular pathway, gene, amino acid or even nucleotide. This in turn has raised the question of the extent to which evolution is predictable. Under similar environmental conditions, is it inevitable that particular phenotypes will evolve time and again? Are such convergent phenotypes also reliant upon convergence at the level of the molecular pathway, gene, amino acid or nucleotide? Or is evolution stochastic in nature; are there many possible ways to make a phenotype?

If, as S.J. Gould put it, “replaying life’s tape” (Gould 1989) was possible then would “the repetition look at all like the original”? Gould argued that as a result of contingency, the evolutionary trajectory taken could substantially differ every time the tape was replayed. Here, “contingency” refers to the unpredictable nature of evolution; just because a particular form exists under the same environment when the tape is replayed doesn’t mean that the outcome will be the same. Furthermore the evolution of a certain form is contingent upon the existence of a series of antecedent forms (Beatty 2006), which won’t necessarily re-occur in the appropriate sequence. However, other evolutionary biologists

have argued that the evolutionary course taken when the tape is replayed may, at least to a certain extent, be predictable (Morris 2010; Stern & Orgogozo 2008; Stern & Orgogozo 2009).

One way in which research has attempted to answer this question of evolutionary predictability is by looking at examples of convergent evolution, preferably in systems in which the fitness benefits have already been established, and by then comparing the loci that determine the adaptive trait in question. This is the approach taken in this thesis in chapter five, where the loci controlling melanism in two Lepidoptera species are investigated and compared to loci found to control melanism in other Lepidoptera species. If genetic parallelism is found to control convergent phenotypes, then this indicates that there may be a certain degree of evolutionary constraint.

Melanism has famously evolved in Lepidoptera as an adaptive response to environmental change. The evolution of 'industrial melanism' in the Peppered Moth *Biston betularia* in response to pollution is the archetypical example of evolution in action (Majerus, 1998). Similarly to the peppered moth, industrial melanism also evolved in other moth species including the scalloped hazel *Odonoptera bidentata* (Bishop et al. 1978; Bishop & Harper 1970), pale brindled beauty *Phigalia pilosaria* (Lees, 1971), mottled umber *Eranis defoliaria*, and dotted border *Agriopis marginaria* amongst others (Majerus 1998).

The locus that controls melanism in *B. betularia* has been mapped to a region of Lepidoptera chromosome 17 (van't Hof et al. 2011). Here, a locus has been identified that controls wing patterning variants in various *Heliconius* species (Baxter et al. 2008; Chamberlain et al. 2009; Ferguson et al. 2010; Joron et al. 2011; Joron et al. 2006; Kronforst et al. 2006; Merrill et al. 2011; Papa et al. 2008), the big eye phenotype in *Bicyclus anynana* (Saenko et al. 2010) and melanism in the peppered moth *B. betularia* (van't Hof et al. 2011), and also in a melanic strain of *Bombyx mori* (Ito et al. 2016). It could be considered that this region is a genomic 'hotspot' for the evolution of adaptive wing patterning in the Lepidoptera.

The concept of adaptive hotspots within the genome is inextricably linked to the idea that evolution may be predictable. The hypothesis postulates that mutations that are able to create the phenotypic diversity upon which selection acts are confined to regions of the

genome in which they can actually cause a change in phenotype, without having pleiotropic effects that limit fitness. If hotspots can be identified then it could be possible to make predictions regarding the nature of adaptation.

Study system: Melanism in Lepidoptera

In this thesis, the melanism in Lepidoptera study system is utilised to further investigate the role that the genomic region described above plays in the evolution of adaptive wing patterning. To this end, this region is assessed to see if it controls melanic forms in two Lepidoptera species that exhibit melanic forms: the pale bridled beauty *Phigalia pilosaria*, and in three melanic strains of the Mediterranean flour moth *Ephestia kühniella*.

The rationale behind choosing distantly related Lepidoptera and also of looking at multiple strains of a single species is that it permits both interspecific comparison – how conserved is the wing patterning hotspot, does it exist in all Lepidoptera? – and intraspecific comparison – are there different ways to create similar phenotypes even in closely related strains of the same species? If phenotypic parallelism at the phenotypic level is frequently accompanied by convergence at other levels then it could indicate that there is some degree of constraint in the areas of a genome in which 'evolutionary relevant' adaptations can occur. Hence, the convergent evolution of melanism amongst Lepidoptera affords the opportunity to investigate the extent to which convergently evolved traits have a similar, or distinct, genetic basis.

Phigalia pilosaria

The pale brindled beauty *Phigalia pilosaria* (Denis & Schiffermüller, 1775), also referred to as *Apocheima pilosaria* (Denis & Schiffermüller, 1775), *Geometra pilosaria* (Denis & Schiffermüller, 1775), *Phigalia pedaria* (Fabricius, 1787). A member of the Family Geometridae, *P. pilosaria* is commonly found across the UK. It is univoltine with adults emerging from January through to March. Females are brachypterous.

Melanic forms of *P. pilosaria* commonly occur in. *P. pilosaria,* which is a well-established example of industrial melanism (Lees 1971; Majerus 1998). There are two melanic forms of *P. pilosaria* associated with industrial melanism, which are controlled by a single locus: f.

monacharia Staudinger is completely dominant to the somewhat lighter and patterned 'intermediate' form, which is itself completely dominant over the recessive non-melanic typical form (Lees 1974).

As an example of industrial melanism, the *monacharia* and intermediate morphs are at higher frequencies in urban and industrial environments than rural areas (Lees 1971). Variation in the frequency of melanic morphs is associated with oak tree reflectance and smoke levels (Lees et al. 1973) with the *monacharia* and intermediate morphs being found at higher frequencies in urban and industrial environments than in rural areas (Lees 1971). Frequency levels of melanic morphs was over 50% in industrial England but started to decline in the 1980s so that by the turn of the 21st Century the frequency of melanic morphs had halved in urban areas (Cook et al. 2002; Lees 1971).

Not all melanic forms of *P. pilosaria* are associated with industrial melanism with other melanic forms having been described (Bretschneider 1939). Melanic forms are known to have existed in the Scottish Highlands prior to the Industrial Revolution (White 1876), where they have continued to persist at frequencies that were little affected by industrial pollution, an occurrence repeated in East Anglia and South Wales (Lees 1971).

Ephestia kühniella

The Indian Meal Moth *Ephestia kühniella* (Zeller, 1879), also sometimes referred to as *Anagasta kühniella* is a member of the Pyralidae family. It is a grain feeder and as such is considered a stored products pest. It is adapted to a domesticated habitat. It lives indoors and is commonly found in bakeries and mills, which makes it amenable for use as a laboratory model. It has now been established as a model for use in cytogenic e.g. (Marec et al. 2001; Marec & Traut 1994; Sahara et al. 1999; Traut & Marec 1996; Traut & Marec 1997; Traut et al. 2007) and pest control studies (Marec 1990; Marec 1991; Rahman et al. 2004). A number of polymorphisms in *E. kühniella* have been described (Caspari 1975) and some of these strains have been maintained in laboratory conditions. Three melanic strains will be used in this study to further investigate the loci that control melanism in the Lepidoptera. These strains are the *Ala nigra* strain, an autosomal dominant melanic mutant (Cotter 1974); *Mel*, an autosomal recessive melanic mutant (Marec 1991); and *abt*, which is also an autosomal recessive mutant (Kuhn & Henke 1929).

These two systems complement each other. *P. pilosaria* represents a 'natural' example of the evolution of melanism. Under the convergent hypothesis, the locus selected in this case is expected to be on chromosome 17. Melanism in *E. kühniella* in contrast, represents a laboratory mutation. Should mutations in laboratory culture occur in a wider set of genes than those represented in natural evolutionary events, then the inference can be made that the convergent pattern in natural systems is associated with reduced pleiotropic effects of mutations at those loci over other potential mutations of apparently similar phenotypic effect (i.e. fewer non-target deleterious impacts). If, in contrast, laboratory mutations typically map to the same region as naturally selected variants, convergence is a property of narrow mutational possibilities.

Outline of Thesis

This thesis investigates the genetic basis of adaptation in two study systems. Chapters two to four are concerned with the evolution of suppression of *Wolbachia*-induced male-killing in the butterfly *H. bolina*. Chapter five investigates whether parallel evolution of melanism in Lepidoptera results from genetic parallelism.

Chapter 2 expands upon previous work investigating the loci involved in the evolution of male survival in populations of *H. bolina*, infected with a male-killing strain of *Wolbachia*. Whilst evolutionary change in a single genomic region has been identified as being necessary for male survival, it is not clear whether this is the only locus involved in the evolution of male-killing suppression, or if secondary suppressor loci exist elsewhere in the genome. This chapter determines whether the suppression of *Wolbachia*-induced male-killing evolved due to changes at a single locus or at multiple loci.

Chapter 3 investigates the sex determination gene *doublesex* as a candidate for the suppression of *Wolbachia*-induced male-killing in *H. bolina.* The chapter investigates *doublesex* as a candidate gene for the evolution of suppression due to it being located in the region syntenic to the suppressor region in other Lepidoptera, its role in the insect sex determination pathway, and altered expression in male-killed strains of the moth *Ostrinia scapulalis*. First, the genomic region in which the suppressor is located is narrowed down to

reduce the number of candidate genes. Next, *doublesex* is mapped to see whether it co-segregates with male survival. A *doublesex* marker is then assessed in wild *H. bolina* samples from populations in which the suppressor is present and compared to those in which the suppressor is absent. Finally, a hypothesis of the molecular changes that enable male survival is proposed.

Chapter 4 documents a selective sweep as it progresses through both space and time. The *H. bolina* of Independent Samoa are infected with a male-killing strain of *Wolbachia*. *Wolbachia* is highly prevalent amongst the *H. bolina* of Upolu and *Savai'i*, the two main islands that constitute Independent Samoa. This resulted in an extremely female-biased sex ratio, which is known to have existed on Indendent Samoa for over a century. In 2005 a male-rescuing suppressor emerged on the island of Upolu, meaning that infected females were able to produce viable male and female progeny. Selection for the suppressor was incredibly strong. A selective sweep for the suppressor occurred spreading firstly across Upolu and then onto Savai'i. The selective sweep that occurred on Upolu has been analysed revealing that this intensive and rapid selection event left a very broad and intense signature of selection on the genomes of Samoan butterflies. This chapter expands upon this analysis of the selective sweep as it progresses in time by investigating its spread from Upolu and across Savai'i. This investigation will assess how the intensity and breadth of the sweep changed as it moved in space, the effect that the sweep had on genetic diversity, and what it reveals about the nature of the locus that controls suppression.

Chapter 5 moves on to explore the second study system within this thesis: parallel evolution of melanism in Lepidoptera. The convergent evolution of adaptive traits, such as melanism in the Lepidoptera, is a common occurrence in nature. A question that evolutionary biologists are currently trying to address is the extent to which phenotypic parallelism arise as a result of genetic parallelism. The locus that controls melanism in the peppered moth has been identified. It has been noted that adaptive wing patterning in other Lepidoptera species has also been mapped to this same genomic region, which is now being promoted as an example of an evolutionary hotspot for adaptive evolution. This chapter maps the locus that controls industrial melanism in the pale brindled beauty *Phigalia pilosaria* to see whether it too is controlled by this same wing patterning hotspot. It then goes on to assess

whether this hotpsot also controls melanism in three strains of the Mediterranean flour moth *Ephestia kühniella*.

This thesis concludes in chapter 6 with a general discussion that commences by summarising the results of the above studies. These results are then synthesised within a wider evolutionary framework. The extent to which the results of this thesis move forward our knowledge of the two study systems and how that goes on to inform understanding of the loci of evolution, the question of whether phenotypic parallelism is accompanied by genetic parallelism, and whether parasites are able to drive the evolution of sex determination systems are discussed. The chapter closes by considering future avenues of work that could extend the results of this study before drawing final conclusions.

Chapter 2

A Single Locus is Necessary for the Suppression of Male-Killing

Abstract

*Hypolimnas bolin*a has evolved the ability to suppress *Wolbachia*-induced male-killing. Infected females that carry this 'suppressor' mutation are able to produce viable male and female offspring. A region of *H. bolina* chromosome 25 has previously been identified as being necessary for the survival of male progeny in suppressor-carrying individuals. An additional 7 autosomes have been assessed and deemed not to contain a secondary locus, required in the suppression of male-killing. However, it is not known whether the region on chromosome 25 is the only region required for male survival or whether there is a secondary suppressor locus elsewhere in the genome of suppressed *H. bolina*. This Chapter demonstrates that there is no evidence that a secondary suppressor locus exists on the sex chromosomes. It also excludes the presence of a secondary suppressor locus in all outstanding autosomes. This shows that suppression of male-killing in *H. bolina* evolved due to changes in a single locus, located on chromosome 25.

Introduction

Sex-ratio-distorting selfish genetic elements, such as male-killers and meiotic drive are common in insects (Duron et al. 2008; Hilgenboecker et al. 2008; Jiggins et al. 2001; Jiggins et al. 1999; Lindholm et al. 2016), and can produce profound population sex ratio bias (Dyson & Hurst 2004). The combination of direct impact on male hosts/spermatogenesis, and Fisherian selection for the production of the rare sex, creates strong selection in the host to counteract and suppress the sex-ratio distorting activity (Cosmides & Tooby 1981; Hamilton 1967; Jaenike 2007a). If a host does successfully manage to evolve the ability to suppress such activity then it renders the sex ratio distorting phenotype cryptic, being revealed only when the element is placed on a new, unsuppressed genetic background (Hornett et al. 2008; Jaenike 2007b). As such, whilst there are few known examples of suppression activity in natural populations, suppression may nevertheless have evolved on many occasions. Those examples where we do see suppression are ones where the suppressor is polymorphic within a population or species (Jaenike 2007b; Majerus & Majerus 2010a), uncovered through introgression when crossed/transinfected to a different population or species (Fujii et al. 2001; Jaenike 2007b; Sasaki et al. 2002; Tao et al. 2001), or is observed to evolve in real time (Charlat et al. 2007b; Hornett et al. 2006).

The genetic architecture of suppression will determine the dynamics of the suppressor. To date, the genetic basis of suppression has been revealed in a number of cases for sex chromosome meiotic drive, cytoplasmic male sterility induced by mitochondria, feminization by *Wolbachia*, and male-killing. In these systems, reviewed below, rescue may be polymorphic or fixed within a population, Y-linked or autosomal, partial or complete, dominant or recessive, single or multiple locus.

a) Meiotic drive: *Drosophila simulans* sex chromosome drive has been widely studied with respect to both the source of drive and the architecture of suppression. Three different sex-ratio meiotic drive suppression systems are known to occur: Paris, Durham, and Winters. In the Winters system, a pair of inversions in the *not much yang* (*nmy*) gene are able to suppress the X-chromosome linked sex ratio distorter, most likely through RNAi activity (Tao et al. 2007a; Tao et al. 2007b). In the Durham sex-ratio distortion system the dominant, autosomal *too much yin (tmy)* locus suppresses an X-linked sex ratio distorter (Tao et al.

2001). In contrast, suppression in the Paris system is polygenic (Mercot et al. 1995) with both Y-linked partial suppression and autosomal-linked complete suppression being observed (Cazemajor et al. 1997).

b) Feminization: The woodlouse *Armadillidium vulgare* carries two different feminising elements, which have both given rise to suppression systems that counteract their effect. Polygenic, autosomal resistance (R) genes are able to prevent transmission of *Wolbachia* onto the next generation, rescuing genetic ZZ males that would otherwise be feminised (Rigaud & Juchault 1992). Another system involves the suppression of the *Wolbachia*-derived feminising *f* element, which has been incorporated into the *A. vulgare* genome and again feminises males. In this system an autosomal, dominant masculinising *M* gene is able to rescue males from the *f* elements feminising effects (Rigaud & Juchault 1993).

c) Mitochondrial-encoded cytoplasmic male sterility (CMS). In CMS, particular hermaphrodite plants are prevented from producing fertile pollen, producing a population that is gynodiocious. CMS is often caused by abnormal mitochondrial open reading frames (ORFs). Fertility can be restored by nuclear-encoded fertility restorer (*Rfi* genes). Although there are many instances of CMS/*Rf* systems, male sterility is often suppressed by pentatricopeptide repeat (PPR) protein genes (Chase & Gabay-Laughnan 2004; Hanson & Bentolila 2004). CMS suppression in rice *Oryza sativa* was found to be polygenic and controlled by the Rf1a gene, and Rf1b genes. Rf1a is epistatic over Rf1b, they use different methods of mRNA silencing to block the protein created by the aberrant CMS causing ORF (Wang et al. 2006).

d) Male-killing bacteria. The ladybird *Cheilomenes sexmaculata* has been found to be polymorphic for a dominant, autosomal single locus that suppresses male-killing (Majerus & Majerus 2010b). The moth *Cadra cautella* has the ability to suppress *Wolbachia*-induced male-killing. This trait is fixed in *C. cautella* but male-killing ability is revealed following transinfection to *E. kühniella* (Sakamoto et al. 2005; Sasaki et al. 2002). There is also evidence for the suppression of sex-ratio distorting elements in a number of Drosophila species. *Drosophila subquinaria* is polymorphic for a dominant, multi locus suppressor of *Wolbachia*-induced male-killing (Jaenike 1996; Jaenike 1999; Jaenike 2007b). A recessive,

autosomal locus is thought to suppress male-killing in *Drosophila prosaltans* (Cavalcanti et al. 1957).

This chapter examines whether suppression of male-killing in *Hypolimnas bolina* is controlled by changes to one or more chromosomes. *Hypolimnas bolina* populations vary in whether male-killing is expressed; infected males (carry suppressor) are commonly found in South-East Asian populations but in only a few Polynesian populations (Charlat et al. 2005; Dyson & Hurst 2004). The presence/absence of suppression is in flux as the suppressor is in the process of spreading across the *H. bolina* range. The oldest evidence of the suppressor was found in Filipino samples dating back to 1890. It was then observed to be present in samples from Thailand and Borneo from the 1970s. When the suppressor does arrive in a new population then it can causes rapid changes in population sex ratios (Charlat et al. 2007b; Hornett et al. 2009). This was observed to occur rather dramatically when the suppressor spread to Independent Samoa, returning the population sex ratio from 100 females per male in 2001 to near unity in 2005 (Charlat et al. 2007b).

We know that the *H. bolina* suppressor is autosomal, dominant, zygotically-acting, and able to restore 1:1 sex ratios (Hornett et al. 2006). A locus, necessary for suppression, has been localised to an approximately 10cM region of chromosome 25 (Hornett et al. 2014). However, the exact nature and molecular mechanism of the suppressor remains unknown. It is also not known whether male rescue suppression is solely determined by the genomic region identified on chromosome 25. This locus is necessary for survival, as established by its presence in all surviving sons. However, it is unclear if it is sufficient, or requires other loci within the genome to achieve male rescue.

Previous work has suggested the locus may act in isolation. Firstly, Hornett *et al*., (2014) excluded an additional 7 chromosomes (5, 6, 8, 9, 15, 17, 18) from being involved in suppression. Secondly, a single locus of suppression is compatible with models for the evolution and spread of the suppressor (Hornett et al. 2010; Hornett et al. 2014). However, there is some evidence indicative of the involvement of other loci in suppression.

Secondary or enhancer loci may also be involved in the suppression of male-killing. Despite there being strong selection for the evolution of suppression it took a long time to evolve. Furthermore, the suppressor appears to have only evolved once and spread, rather than

having evolved multiple times *de novo* (evidence for this is presented in chapter 3). As selection pressure for the evolution of the suppressor was so strong and took a relatively long time to occur the suppressor must be hard to evolve. The evolution of suppression must therefore be more complex than an amino acid or base pair substitution; rather a restructuring event e.g. inversion, duplication event or contingent on changes at multiple loci, is more likely to have occurred. Models predict that when the suppressor is introduced into a population then the locus of suppression should become fixed within the population and that linked alleles at nearby loci should hitchhike to high frequencies (Hornett et al. 2010). However, analysis of the genomic impact of the suppressor into the *H. bolina* population onto the island of Upolu, Independent Samoa found that allele frequency changes were lower than expected, which may indicate the involvement of further loci involved in suppression (Hornett et al. 2014).

More evidence that the suppressor may potentially be a multilocus trait arise from there being some discordance between male survival rates and a fully penetrant male-killer. For a single locus model, a heterozygous individual carrying the suppressor should produce two females per male if rescue is fully penetrant. However, sex ratios of four females per male were observed both in Hornett *et al*, 2006 and in Hornett *et al* 2014. This implies either multiple loci, or incomplete penetrance of rescue.

In this chapter markers are developed to complete coverage of all 31 chromosomes in the *H. bolina* genome, and analyse (using the progeny from crosses in Hornett *et al*., 2014) whether suppression also requires the involvement of loci from chromosomes hitherto untested. This would be determined by all male progeny from a cross carrying one of a pair of maternal haplotypes for a particular chromosome.

Methods

Mapping Families

Mapping was carried out using the mapping families created by, and described in Hornett *et al*., (2014) (figure 2.1). These crosses were of a reciprocal backcross design. A *Wolbachia*-infected female homozygous for the suppressor (SS) was crossed to a *Wolbachia*-infected male lacking the suppressor (ss). The female originated from the Philippines, where the suppressor is ubiquitous, whilst the male was derived from Moorea, French Polynesia where the suppressor is not present. The F_1 females resulting from this cross were heterozygous for the suppressor (Ss). To create the female-informative family, these females were backcrossed with Moorean males lacking the suppressor (ss). As female Lepidoptera do not undergo recombination (Suomalai.E et al. 1973; Turner & Sheppard 1975) the female-informative family can be used to determine whether any particular linkage group is necessary for suppression (marker would show Mendelian segregation in F_2 females and be present in all F_2 males). The reciprocal, male-informative family was created by crossing an F_1 male heterozygous for the suppressor (Ss) with Moorean females lacking the suppressor (ss). As recombination does occur in male Lepidoptera, the male-informative families can be used to map traits within a linkage group.

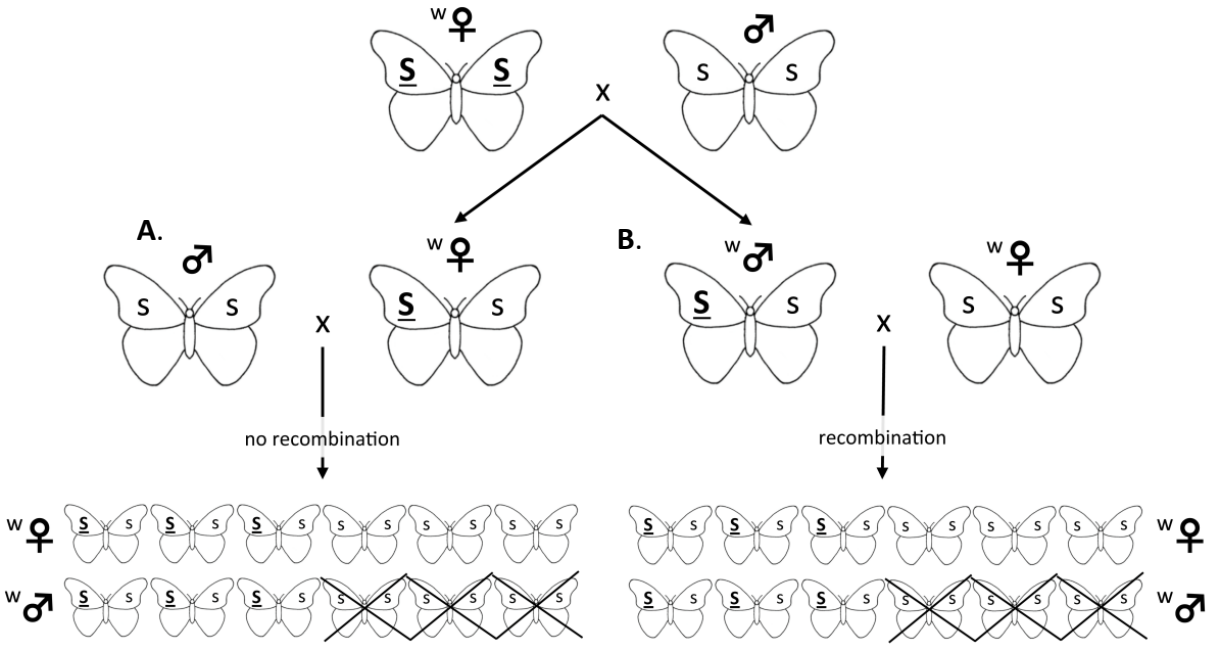

Figure 2.1: Crossing scheme used to generate *Hypolimnas bolina* mapping families: A *Wolbachia*-positive (w) female homozygous for the suppressor (SS) was crossed with an uninfected male lacking the suppressor (ss). A. shows the female-informative family, which was created by backcrossing *Wolbachia*-positive, heterozygous (Ss) daughters with uninfected, unsuppressed males. B. shows the male-informative family which was created by crossing *Wolbachia*-positive, heterozygous (Ss) sons from the initial pairing with infected females lacking the suppressor. In the F2 progeny, half of all males in both the male- and female-informative families would lack the suppressor and thus suffer from male-killing. (Figure reproduced from Hornett *et al*., 2014).

Marker development

A first pass genome assembly was used as a basis for marker design. DNA was extracted from the abdomen of a female butterfly of Thai origin (supplied from London Pupae Supplies) following a Phenol-Chloroform protocol (appendix i).

This DNA was used to create two 100bp Illumina paired end libraries: TruSeq fragment library and Nextera large-insert library. These libraries underwent paired-end sequencing on the Illumina HiSeq platform. The resulting data is summarised in table 2.1**.** Library creation, sequencing and trimming of the subsequent sequence data was carried out by the Centre for Genomic Research, University of Liverpool. Genome size and coverage was estimated via Kmer frequency spectra analysis (k=21), which revealed a rather heterozygous genome with an estimated genome size of 750Mbp with 40 x coverage.

Library	Number of Reads (bp)	Mean Read Length (bp)	Read Length Standard Deviation (bp)	Mean Base Quality Score	GC content %
Fragment Library R1	92 790 230	86.9	23.8	36.7	29.1
Fragment Library R2	92 790 230	92.7	18.4	37.3	29.1
Large-Insert Library R1	87 373 709	98.9	6.5	38.4	32.5
Large-Insert Library R2	87 373 709	97.6	10.9	37.8	32.5

Table 2.1: Summary of trimmed fastq sequence data files resulting from Illumina sequencing of *H. bolina* genome.

A *de novo* first pass assembly was created using CLC Genomics Server. Both fragment and large-insert libraries were used to create fragments from reads and also to optimise the graph based on paired reads. The default word size was used, a default maximum bubble size of 50bp, and default minimum contig size of 200bp were also used. The resulting assembly consisted of 561 288 contigs with a range of 200 – 50 189 bp, and an N50 of 1374bp. Read mapping of the assembly was carried out using Bowtie2 v. 2.2.1 in local alignment mode, giving an overall alignment rate of 91.5%. The assembly was used to create a custom, local nr BLAST database in Geneious Pro version 5.6.6 (Kearse et al. 2012).

Owing to the high level of gene order synteny across the Lepidoptera, (d'Alencon et al. 2010; Pringle et al. 2007), potential genomic regions were selected to cover outstanding *H. bolina* chromosomes based on their location and non-paralogous nature, in the *Bombyx mori* genome. *B. mori* coding sequence of the targeted genes was acquired from SilkDB (Duan et al. 2010; Wang et al. 2005) and used to perform a local tBLASTx on the *H. bolina* database. Contigs from genes that resulted in a single hit were retrieved and aligned to *B. mori*, in order to infer intron/exon boundaries. Primers were designed to be located in exonic regions, so as to amplify introns. Primers were designed with the aid of Oligo6 and

ordered from Eurofins. Marker loci information is described in table 2.2. PCR conditions were optimised using test *H. bolina* DNA.

Marker amplification and sequencing

PCR was performed in a total volume of 12 µl consisting of 4 µl ddH2O, 6µl of 2x GoTaq Hot Start Green Master Mix (Promega), 0.5pmol of forward primer, 0.5pmol reverse primer and 1 µl of DNA. PCR cycling conditions consisted of an initial denaturation period of 1 minute at 95°C, followed by 35 cycles of 15 seconds at 95°C, 15 seconds at annealing temperature, extension time at 72°C, this was followed by a final extension of 7 minutes at 72°C. The annealing temperature and extension time for each marker can be found in table 2.3

Resulting PCR products were visualised on an agarose gel. Samples that had successfully been amplified were then purified and sequenced using Sanger technology following the PCR product and sequencing protocol (appendix ii).

Chromosome	Marker Name	Gene	*B. mori* Orthologue	*B. mori* Accession Number	Exon/Intron number
Z	Kettin	Kettin	BGIBMGA000622	NM_001114876.1	intron 3
2	9724	Putative ammonium transporter 2	BGIBMGA009724	XM_004933286	intron 8
3	Sanna	Protein msta, isoform B	BGIBMGA008839	XM_012691453	intron 5
4	3145	Piezo-type mechanosensitive ion channel	BGIBMGA003145	XM_012694483	intron 15
5	IQGAP1	Ras GTPase-activating-like protein IQGAP1	BGIBMGA003700	XM_012693848	intron 6
6	Hb60	Ribosomal protein L15	BGIBMGA006414	NM_001043697	introns 2 and 3
7	Pixie	Ribonuclease L inhibitor homolog (Rli)	BGIBMGA010129	NM_001043446	intron 2
8	Hb46	Ubiquitin-conjugating enzyme E2	BGIBMGA005425	NM_001046942	introns 1 and 2
9	Hb111	Male reproductive organ angiotensin-converting enzyme-related protein 1	BGIBMGA002526	AB485780	exon 8/intron 8
10	Laccase 2	Laccase 2A	BGIBMGA006745	AB379590	intron 4
11	12078	Sorting nexin-14-like	BGIBMGA012078	XM_012696095	intron 1
12	Hb68	Ribosomal protein L9	BGIBMGA010571	NM_001043679	intron 1
13	Yellow	Yellow 7	BGIBMGA001149	JN977509.1	intron 2
14	Hb 113	Ribosomal Protein S3 (RpS3)	BGIBMGA009319	NM_001043788	intron 2
15	Hb35	Rb protein L5	BGIBMGA007879	AY769272	intron 5
16	Nip Snap	NIPSNAP protein	BGIBMGA012859	NM_001046674	intron 3
17	Hb66	Ribosomal protein P40c	BGIBMGA007311	AB062685	intron 4
18	Hb115	Elongation factor 1 gamma	BGIBMGA008302	NM_001043387	intron 9
19	dhpr	Dihydropteridine reductase	BGIBMGA003987	XM_004922396	exon 1
20	CG6702	Ccalbindin-32-like	BGIBMGA004330	XM_012689715	intron 3

Chromosome	Marker Name	Gene	*B. mori* Orthologue	*B. mori* Accession Number	Exon/Intron number
21	mapkin erkin	MAP kinse-ERK kinase (Mek)	BGIBMGA001458	XM_012693302	intron 3
22	366	Cation-transporting ATPase 13A3	BGIBMGA000366	XM_012695359	intron 6
23	11189	Synaptotagmin-7	BGIBMGA011189	XM_012694295	intron 2
24	Ferritin HC	Ferritin Heavy Chain ferritin subunit precursor	BGIBMGA008768	DQ898220	intron 1
25	5100	Uncharacterised	BGIBMGA005100	XM_012689849	intron 1
26	10725 F2R2	Coatomer protein complex subunit alpha (COPA)	BGIBMGA010725	NM_001172721	intron 12
27	Cubitus Interruptus	Transcriptional activator cubitus interruptus	BGIBMGA004545	XM_012697765	intron 2
28	Fatbody	Fatbody protein 3Rev-G1	BGIBMGA013782	NM_001195451	exon 1
29	Black	Black / aspartate decarboxylase	BGIBMGA012088	KM523624	intron 3
30	Acyl co Red	Fatty-acyl CoA reductase	BGIBMGA011148	XM_004925935	introns 6 and 7
31	Puff	Puff-specific protein Bx42	BGIBMGA008202	XM_004923701	intron 2

Table 2.2: Loci information for markers covering all autosomes and Z chromosome.

Chromosome	Marker	F Primer (5'>3')	R Primer (5'>3')	F/R	Ta (°C)	Extension Time (sec)	Reference
Z	Kettin	GAACTTGGCGAAGCTGTTACGTC	GCCGCAAATCCAAAATCATAAG	R	60	30	This study
2	9724	GCATCAGCAGTTCACGGAGC	CAGCACGAGGTCAATAAGCCA	F	57	60	This study
3	Sanna	CGAGTTCGGAACCATGT	GCAAGCATCAGGATCTCCCTT	R	53	45	This study
4	3145	GGCCACGTCCAAGTACCTG	TCGGAGGGTTCGGTTGTC	F	57	60	This study
5	IQGAP1	TGACCTGTACAGAGCTTTGACGAG	GTGTGACAAACAATCCAGGGAC	F	60	60	This study
6	Hb60	CATGGAGTCAACCAGCTGAA	CCACCCTTTGTCTGGGAGTA	F	53	45	Hornett *et al.*, (2014)
7	Pixie	AGAGCCACGGAAATGAGACAA	TGTCCCTTTGATGCCATTACC	R	60	45	This study
8	Hb46	GATTGGCCGAAGAGAGAAAA	AGGGTACACATTGGGATGGA	F	53	45	Hornett *et al.*, (2014)
9	Hb111	TCCTACCATTCGCGTACACA	ACTCACCTCGCGTATTCAACATC	R	53	15	This study
10	Laccase 2	AAACGAGCCCAACAAC	GTCAGCTGGTCTGTTGTCACC	R	57	90	This study
11	12078	CAGGGTTGTCATGCATGTGG	AGCTGTTGGCGCAGTTCATC	F	56	45	This study
12	Hb68	CTGTGGACATTCGCATGGTA	TTAATGGGGAAATGGGCATA	F	53	30	Hornett *et al.*, (2014)
13	Yellow	AATGCGACCGTCTTTG	CAGTTCCTTGAGTGAGCTTGTCC	R	53	75	This study
14	Hb 113	AGTTGAGCTTTACGCCGAGA	ACGCTGGCCTCTCAGTTTAC	R	57	60	Hornett *et al.*, (2014)
15	Hb35	TTGGAGCTATGAAGGGTGCT	TGCCGAACTGTCTCTTGTAGG	F	53	30	Hornett *et al.*, (2014)
16	Nip Snap	GTGGGCGAGAACTCAGT	ATCGCAAATCGGAAGGC	F	43	45	This study
17	Hb66	GTGAGGTTCTTCGCTTGAGG	ATGAGGTGACGGGTTCCAC	F	53	45	Hornett *et al.*, (2014)
18	Hb115	GACGACGACTCCACCATCTC	CCGGACCACGAGAAGTAGTC	F	53	60	Hornett *et al.*, (2014)
19	dhpr	GACTACCGCTCGGAGG	GCAGCTTGCGTCAATCATTTC	F	53	45	This study
20	CG6702	ACTGAAGGAAGCGAAGAA	AGATAACTGGAGTCGGCCATC	R	57	45	This study

Chromosome	Marker	F Primer (5'>3')	R Primer (5'>3')	F/R	Ta (°C)	Extension Time (sec)	Reference
21	Mapkin Erkin	TGCTTATCGCGCAAATA	TAATGAAAGTTCGCCACAAGTCC	F	53	60	This study
22	366	GTATCGCCGCTATCGGAATG	AGATGCGGTTGGCCTTGA	R	60	30	This study
23	11189	AAACACTTCGCCGATGGAG	AGCCAAGGATTTAAGCGGGAC	F	52	60	This study
24	Ferritin HC	TGACGAATGGCGTTCCATC	GGCCGCGTCCATCCA	F	55	120	This study
25	5100	GGAGCATGCCCTTAAGTAAGTTTC	CGGGTCCGTTCGTTATCTG	R	55	45	This study
26	10725	GCCTGATCCTTTCGCTAACATC	AGTGTGGCAACATCGAAGTGG	F	57	60	This study
27	Cubitus Interruptus	GACCCAACGACAGACAA	TGCGGCAGAAAGTACAGTCCA	F	60	30	This study
28	Fatbody	CCCTTGGACAGCCTACA	TTGAATCCGTGGTTCCTTTGA	F	53	30	This study
29	Black	CGTGGACGCGGCTTG	ACCTTGCGGCTGATACGTGA	R	58	45	This study
30	Acyl co Red	ATGATCGACGCTGGAC	TGCTGATACGGCGATGGAC	F	55	75	This study
31	Puff	GTTGAAAGCTAAGCGTGTCGG	TGGCTCCCATACCCAGAGG	F	56	30	This study

Table 2.3: Primer information and PCR conditions for amplification of markers covering autosomes and the Z chromosome.

Ensuring that markers cover all chromosomes:

The *H. bolina* genome consists of 31 chromosomes (Robinson 1971). To ensure that the markers were suitable for use in this study a χ^2 test was performed to assess whether markers showed Mendelian segregation. A marker would be considered to be Mendelian if the ratio of genotypes in female F_2 offspring did not deviate from the expected 1:1 Mendelian ratio.

To check that the markers used did indeed represent all 31 chromosomes, the segregation pattern of the F2 offspring was assessed visually for all autosomal markers to determine if it was different for each marker (figure 1). The marker used for the Z chromosome has been shown to be located on the Z chromosome in other Lepidoptera, and so should be trustworthy to use as a Z chromosome marker.

Determining whether there is a secondary locus necessary for suppression

Autosomes: To determine whether a linkage group contained a mutation that was necessary for male survival markers were first tested in the F_1 parents of the female-informative family. Markers that contained SNPs that were heterozygous in the female parent but homozygous in the male parent were then sequenced in the F_2 offspring, consisting of 15 daughters and 5 sons. A chromosome would be considered to be potentially associated with suppression if the marker was found in all male offspring and showed Mendelian segregation, being present in half, of the female offspring. Markers found to be present in all sons were then assessed in a second female-informative mapping family, to confirm or exclude association with suppression.

Sex chromosomes: It would not be possible to exclude/include the sex chromosomes as being necessary for male survival using the female-informative family. Lepidoptera have a ZW sex determination system in which females are the heterogametic sex. Being hemizygous for each of the sex chromosomes renders it impossible for SNPs to be found in the maternal parent. To overcome this, the male informative family was instead used. A SNP heterozygous in the male parent and hemizygous in the female parent was identified in the paternal parent of the male-informative family. 16 male F_2 offspring from the male-

informative mapping family were sequenced and scored for the male-informative marker. A χ^2 test of association was performed in R (table 3). If the SNP was found to co-segregate with male survival then it would indicate that a mutation on the sex chromosomes was also involved in suppression of male-killing. If the SNP was found not to segregate with male survival then this would suggest that a secondary locus required for male survival is not located on the Z chromosome. It is however possible that a Z-chromosome locus, not in linkage with the tested marker, may still act as a secondary locus of suppression.

Results

Is there a mutation necessary for male survival located on the sex chromosome?

A region of the Z-linked Kettin gene, containing a SNP found to be heterozygous in the male parent and hemizygous in the female parent of the male-informative family, was amplified in 16 male F2 progeny from the male informative family. Of these, 8 individuals were heterozygous and 8 homozygous for the paternal allele. The Z-linked marker did not co-segregate with male survival (χ^2=0, df=1, N=16, p=1).

Is there a secondary locus, necessary for male survival, located on the autosomes?

A chromosome would be identified as being necessary for male survival if a single allele (either homozygous or heterozygous) was present in all male offspring but showed Mendelian segregation, by being present in half of the female offspring. This was the case for the markers representing chromosomes 14 and 25. Hornett *et al*., (2014) previously demonstrated that a region on chromosome 25 is necessary for male survival. The chromosome 14 marker was further investigated by testing it in an additional female-informative family (family b), which was created in the same manner and at the same time as the family used in the rest of this study. In this family the chromosome 14 marker showed Mendelian segregation in F_2 female offspring but was not found to co-segregate with male survival, with 7 male offspring being heterozygous and 9 homozygous for the marker (see 14b, Table 3). It is therefore likely that the co-segregation of the chromosome 14 marker with male survival in the original female-informative mapping family was due to sampling error. No other markers showed co-segregation in male offspring, implicating that there isn't a secondary locus of selection and that suppression of male-killing in *H. bolina* is a single locus trait.

Do markers represent all 31 chromosomes?

The F2 female-informative offspring were amplified and sequenced with 30 autosomal markers. The markers were assessed to ensure they represented all 31 chromosomes and

were suitable to be used in the study by showing Mendelian segregation. All 30 autosomal markers displayed a unique pattern of segregation in the F_2 offspring (figure 2.2), showing that each marker belonged to a different linkage group (which comprises of the entire chromosome in the female-informative mapping family). The Z-chromosome used is a well-established marker for this chromosome. This suggests that all 31 chromosomes were represented in this study. Markers for all chromosomes, save for chromosome 8 were assessed to be Mendelian, based on a X^2 test (table 2.4), making them suitable for use in this study. The segregation pattern of the chromosome 8 marker was found to deviate from the expected Mendelian ratios, however further assessment indicates that this may result from sampling error as when both male and female offspring are considered (male offspring will show Mendelian segregation for chromosomes that are not necessary for suppression, which is the case for chromosome 8) the segregation pattern does not diverge from expected Mendelian ratios (χ^2 =3.200, df=1, N=20, p=0.074).

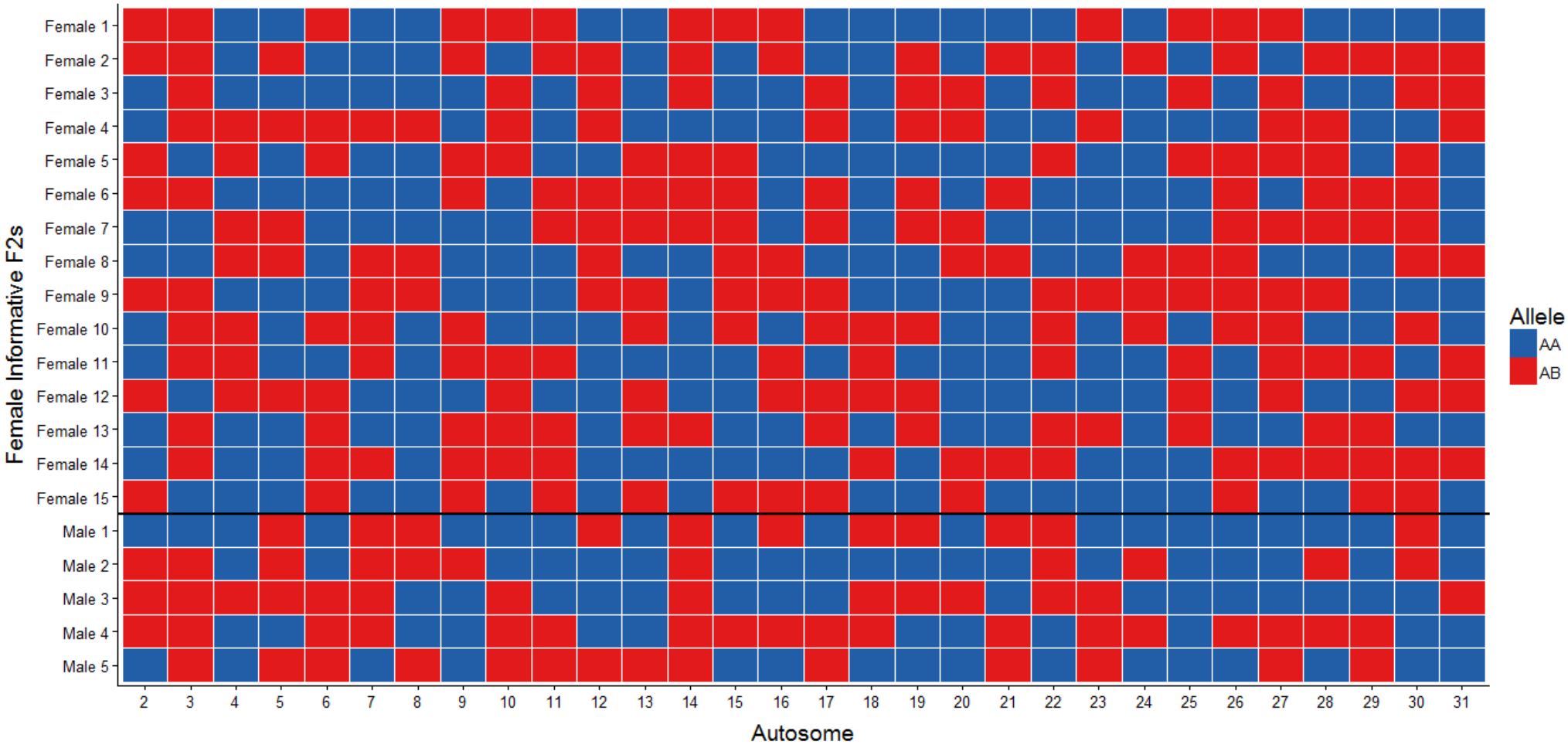

Figure 2.2: Female-Informative F_2 offspring. Map of female-informative mapping family progeny showing segregation pattern of inheritance of a female-informative SNP marker. The two allele variants are colour coded with homozygotes being blue and heterozygotes red. The population, suppressed or unsuppressed, from which each allele is derived, is not known.

Chromosome	AB	AA	X^2	p-value
2	7	8	0.067	0.796
3	10	5	1.667	0.197
4	7	8	0.067	0.796
5	5	10	1.667	0.197
6	8	7	0.067	0.796
7	6	9	0.600	0.439
8	3	12	5.400	0.020
9	9	6	0.600	0.439
10	8	7	0.067	0.796
11	8	7	0.067	0.796
12	7	8	0.067	0.796
13	8	7	0.067	0.796
14	7	8	0.067	0.796
14 (family b)	9	10	0.053	0.819
15	8	7	0.067	0.796
16	7	8	0.067	0.796
17	9	6	0.600	0.437
18	4	11	3.267	0.707
19	8	7	0.067	0.796
20	6	9	0.600	0.439
21	4	11	3.267	0.707
22	8	7	0.067	0.796
23	4	11	3.267	0.707
24	4	11	3.267	0.707
25	8	7	0.067	0.796
26	10	5	1.667	0.197
27	10	5	1.667	0.197
28	9	6	0.600	0.439
29	7	8	0.067	0.796
30	10	5	1.667	0.197
31	7	8	0.067	0.796

Table 2.4: Assessing whether markers show Mendelian segregation: Segregation of parental alleles in female-informative family F_2 females. Number of female F_2 offspring from female-informative family with either heterozygous (AB) or homozygous (AA) allele for markers for chromosomes 2-31. Results of χ^2 test to determine whether markers are Mendelian are also shown.

Discussion

The butterfly *H. bolina* has evolved the ability to suppress *Wolbachia*-induced male-killing. A region of *H. bolina* chromosome 25 has previously been shown to be necessary for male survival in butterflies that carry the suppressor mutation. A further 7 autosomes (5, 6, 8, 9, 15, 17, 18) have also been assessed and were shown not to be required in the suppression of male-killing. It was not however known whether a secondary suppressor locus existed elsewhere in the genome. Models of the evolution of the suppressor are consistent with suppression being a single locus trait (Hornett et al. 2006; Hornett et al. 2010). However, despite there being a very strong selection pressure for the evolution of the suppressor it took a long time to evolve. This indicates that the evolution of suppression is either contingent on changes at multiple loci or an unusual form of mutation (rather than a simple SNP). Further evidence that multiple loci may be involved arises from the fact that the suppressor doesn't rescue all males – this could indicate that multiple loci are involved (all necessary loci may not segregate together, meaning that not all males survive) or that the suppressor is not completely penetrant. This chapter investigated whether the evolution of the suppression of male-killing requires the involvement of any additional loci to that already identified on chromosome 25.

There was no evidence that a locus necessary for suppression of male-killing exists on either of the sex chromosomes. Lepidoptera have a ZW sex chromosome system in which females are the heterogametic sex. The absence of the W chromosome in males, excludes this chromosome from carrying a mutation that males require to survive. No association was found between the Z marker and male survival. This marker was only tested in the male-informative family so, unlike the autosomes, the entire chromosome cannot be excluded based on this single marker. Rather the lack of association shows that neither this marker nor any regions of the chromosome linked to this marker are required for male survival. To fully establish that the Z chromosome does not contain a region required for suppression, the same process would have to be repeated with a number of markers spread along the length of the Z chromosome. Nevertheless, there is no evidence that the sex chromosomes are required for suppression of male-killing.

A secondary locus, necessary for suppression, was not found on any of the autosomes. When taken alongside the lack of evidence for a secondary suppressor locus on the sex chromosomes, this reveals that the evolution of the suppression of male-killing in *H. bolina* is the result of changes at a single locus on chromosome 25.

This study developed markers to cover the Z sex chromosome and the 22 autosomes that had not yet been assessed for the presence of a locus necessary for male survival. The 30 autosomal markers were shown to segregate independently in a Mendelian fashion in the female-informative mapping family. This shows that the markers used in this study are representative of all of the *H. bolina* autosomes. The marker used for the Z chromosome is well-established as being located on the Z chromosome in Lepidoptera.

This study was only able to show that evolutionary changes at a single genomic region are required for the suppression of male-killing. This does not rule out the possibility that other loci may be involved in enhancing the survival chances of males. Evidence from crossing studies implies that the suppressor isn't fully penetrant, in that it does not rescue all male offspring. It is therefore possible that secondary loci may exist that affect the ability of the suppressor.

Future work could investigate the existence of enhancer loci. This could be achieved by performing a genome-wide association study (GWAS) to look for variants associated with male survival. This study did not identify which of the assessed alleles was derived from the suppressed population and which from the unsuppressed population. This could be rectified by sequencing the grandparents of the female-informative family offspring.

This chapter has shown that the previously-identified suppressor region on chromosome 25 is the only one required to suppress *Wolbachia*-induced male-killing in *H. bolina*. Even though a very strong selection pressure existed for the evolution of a suppressor it took a relatively long time to evolve. That genetic changes at only a single locus control suppression suggests that the mutation must have been difficult to evolve; it is unlikely to result from a simple basepair change or amino acid substitution. Rather, it suggests that the suppressor must be more complex, it could involve structural changes such as duplications or inversions. It could also be possible that the suppressor locus is a supergene, a group of tightly-linked loci that act to control an adaptive phenotype, such as that which controls

mimicry in the butterfly *Heliconius numata* (Joron et al. 2011; Joron et al. 2006) or heterostyly in the Chinese primrose *Primula sinensis* (Mather 1950). The nature of the suppressor locus will be explored further in chapter 3.

Chapter 3

doublesex: A candidate gene for suppression?

Abstract

The genetic change responsible for the suppression of *Wolbachia*-induced male-killing in *Hypolimnas bolina* has been localised to an approximately 10cM region of *H. bolina* chromosome 25. This chapter first establishes a wider set of markers in this region, and then utilizes these markers to further narrow this region. This analysis locates the suppressor to a 2.4cM area of chromosome 25. This region is confirmed as containing the candidate gene *doublesex*, which is involved in the sex determination pathway. The chapter goes on to propose the hypothesis that suppression of male-killing in *H. bolina* results from the duplication, or partial duplication, of *doublesex*. Evidence for this hypothesis is threefold. Firstly, *doublesex* is located within the suppressor region and co-segregates with male-survival. Secondly, *doublesex* is located in the part of the suppressor region that shows the greatest signature of selection associated with a selective sweep for the suppressor on Upolu, Independent Samoa. Finally, wild *H. bolina* populations, in which the suppressor is present, show an excess of apparent heterozygotes at the *doublesex* locus, a phenomenon that could be explained by each individual being homozygous for 2 copies of *doublesex*. If this hypothesis is true then the *H. bolina* - *Wolbachia* system acts as a demonstration that parasites can drive the evolution of major developmental pathways. Future work should aim to validate or refute the hypothesis that suppression of *Wolbachia*-induced male-killing in *H. bolina* results from duplication of *doublesex*, and to allow further understanding of the molecular mechanisms by which suppression operates.

Data for the markers Hb80, Hb86, Hb89, Hb94, Hb99, Hb131, Hb136, Hb147, Hb152, Hb175, and Hb176 is derived from Hornett *et al*., 2014 but is shown in this chapter in order to put the data generated in this chapter into context.

Introduction

The butterfly *Hypolimnas bolina* has evolved the ability to suppress the male-killing phenotype of *Wolbachia* infection (Hornett et al. 2006). Suppression is dominant, highly penetrant and zygotically-acting and is known to result from evolutionary change within nuclear genes of the *H. bolina* genome (Hornett et al. 2006). As demonstrated in chapter 2, suppression is a single locus trait, which has been localised to an approximately 10cM region of *H. bolina* chromosome 25 (Hornett et al. 2014). The mechanism by which suppression occurs has yet to be fully characterised.

Hornett *et al*., (2014) note that in other Lepidoptera species the gene *doublesex* is found in a genomic region that is orthologous to the *H. bolina* suppressor region. They hypothesise firstly that due to high levels of gene order synteny within Lepidoptera (d'Alencon et al. 2010; Pringle et al. 2007), *doublesex* is also thus likely to lie within this region in *H. bolina*. They further hypothesise that, owing to its role in the insect sex determination cascade and altered splicing in male-killed individuals in other Lepidoptera (Sugimoto & Ishikawa 2012; Verhulst & van de Zande 2015), that *doublesex* represents a strong candidate gene for suppression.

The development of sex in the Lepidoptera is controlled through the sex determination pathway (figure 3.1), which has been characterised in the silkworm *Bombyx mori* (Fujii & Shimada 2007; Kiuchi et al. 2014; Ohbayashi et al. 2001; Ohbayashi et al. 2002). Like most Lepidoptera, *B. mori* has a ZW sex chromosome system, in which females are the heterogametic sex (ZW) and males homogametic (ZZ)(Traut et al. 2007). The Z chromosome contains the *Masculinizer* gene (*Masc*) and the W chromosome contains a dominant Feminizer *Fem*, a PIWI-interacting RNA (piRNA) precursor. *Masc* codes for a lepidopteran-specific CCCH-tandem zinc finger protein that signals for both dosage compensation of the Z chromosome and for male-specific splicing of *doublesex.* The W locus *Fem* codes for a piRNA that targets and silences *Masc* mRNA through cleavage. This silencing by *Fem* piRNA is required for female-specific splicing of *doublesex* (Kiuchi et al. 2014).

Doublesex is the gene at the bottom of the sex determination cascade; it is the switch that signals whether a cell should develop male or female characteristics. *Doublesex* consists of 6 exons that are alternatively spliced to produce male and female isoforms. In Lepidoptera,

males splice a single male-specific isoform that excludes exons 3 and 4 (Ohbayashi et al. 2001; Shukla & Nagaraju 2010; Sugimoto et al. 2010; Wang et al. 2014). Comparatively, females have been found to express numerous female-specific isoforms, with intron 3 being a hotspot for splicing events (Duan et al. 2013a; Wang et al. 2014).

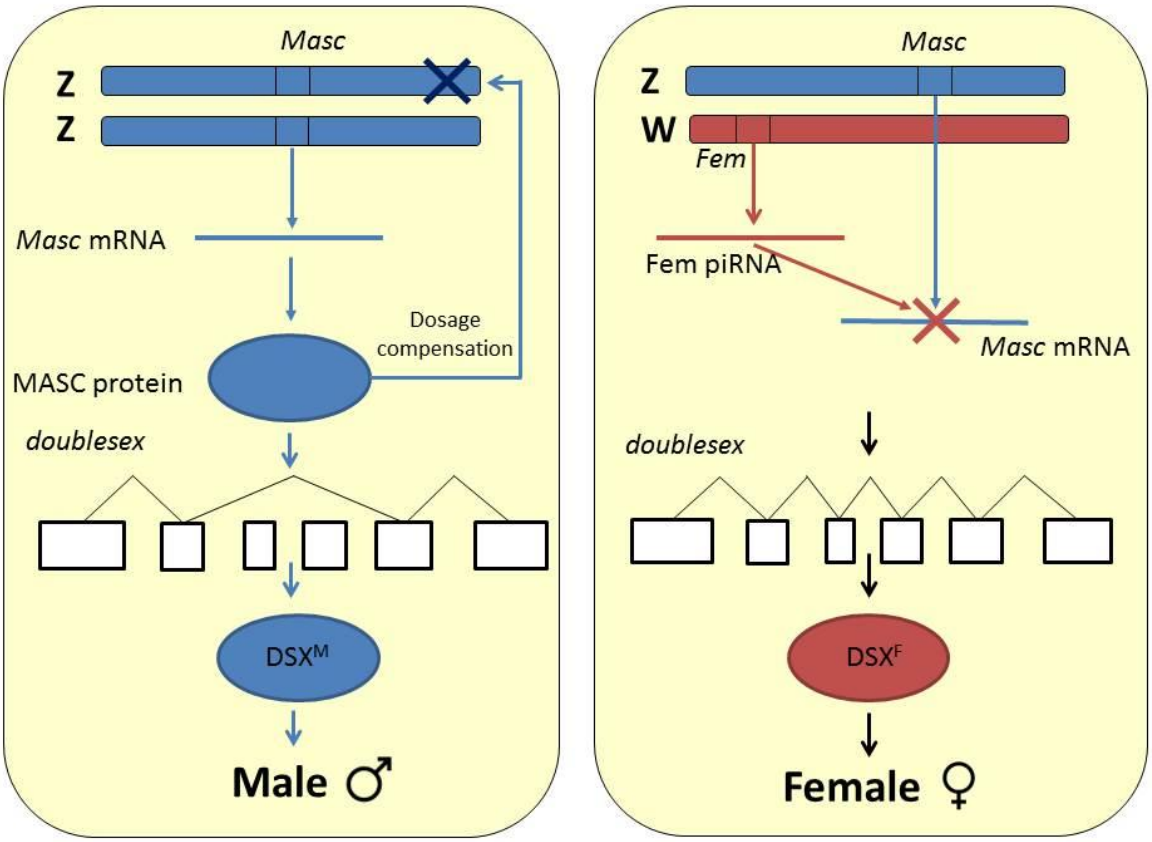

Figure 3.1: Sex-determination pathway, as elucidated in the Lepidoptera *B. mori*. The sex chromosomes lie at the top of the sex determination cascade. In males *Masc* is located on the Z chromosome. It codes for a protein that signals for dosage compensation in the Z chromosomes and for male-specific splicing of *doublesex* to produce DSX^M, the male isoform of *doublesex*, which signals for cells to develop male characteristics. In females a piRNA is encoded by *Fem* on the W chromosome. Fem piRNA targets *Masc* mRNA produced by the female Z chromosome. In the absence of the *Masc* protein *doublesex* alternatively splices to create female splice variants, which codes for DSX^F, a female isoform of *doublesex*. This protein signals for cells to develop female characteristics. Blue arrows show the cascade with a signal originating on a Z chromosome. Red arrows show progression of the cascade with a signal originating on the W chromosome and black arrows the pathway that occurs in the absence of *Masc*.

Wolbachia cause male-killing in Lepidoptera by interfering with its host sex determination pathway. For male development to occur *Masc* is needed to signal dosage compensation of Z-linked genes and also to signal male-specific splicing of *doublesex* (Kiuchi et al. 2014). In uninfected males, *Masc* expression peaks 6 hours after oviposition and then quickly decreases. Comparatively, in *Wolbachia*-infected males there is no such peak; *Masc* expression remains low throughout development (Fukui et al. 2015). Under conditions of low *Masc* levels, the *doublesex* gene alternatively splices to produce female-specific isoforms of *doublesex* (Kiuchi et al. 2014). This process occurs even in males that are genetically male (ZZ) but have reduced *Masc* levels as a result of *Wolbachia* infection. In these cases the male-specific isoform of *doublesex* is not detected (Fukui et al. 2015). In the adzuki bean borer moth *Ostrinia scapulalis* such individuals were found to express female isoforms of *doublesex*. Death occurred in embryos in which there was a discrepancy between genetic and phenotypic sex i.e. ZZ genetic males that expressed female isoforms of *doublesex* (Sugimoto & Ishikawa 2012; Sugimoto et al. 2015).

Death in male-killed embryos resulted from a failure of Z-linked dosage compensation in *B. mori* (Kiuchi et al. 2014) and *O. scapulalis* (Sugimoto et al. 2015). Normally, the presence of *Masc* protein will trigger dosage compensation of Z-linked genes 24-48 hours post oviposition in genetic males (Fukui et al. 2015; Kawamoto et al. 2015). However reduced levels of *Masc* in *Wolbachia*-infected, male embryos mean that this process doesn't occur. Instead expression of Z-linked genes is upregulated. A two-fold increase in expression levels of Z-linked genes results in embryo death. Artificial replenishment of *Masc* in depleted male embryos rescues them from death (Fukui et al. 2015; Sugimoto et al. 2015).

In summary, *Wolbachia* achieves male-killing in Lepidoptera by interfering with the sex determination cascade. Via an as yet unknown factor, *Wolbachia* is able to keep *Masc* levels in infected male embryos low. This has the two-fold effect of triggering female-specific, rather than male-specific splicing of *doublesex,* and also prevents successful dosage compensation of Z-linked genes.

Using linkage mapping, Hornett *et al*., (2014) were able to elucidate an approximately 10cM region of chromosome 25 that was deemed necessary for males to survive *Wolbachia*-induced male-killing. They went on to analyse the genetic impacts that a selective sweep for

suppression had around this locus. The selective sweep in question occurred on the island of Upolu, Independent Samoa. In this population, a stable female-biased sex ratio of 100:1 had been present for some 100 years (Dyson & Hurst 2004). That was until suppression of male-killing arrived on the island in around 2005. Selection for the suppressor was intense and rapidly swept across the island, so much so that a 2006 survey determined that the sex ratio had changed from 100:1 in 2001, when the island was last surveyed, to 1:1 (Charlat et al. 2007b; Dyson & Hurst 2004).

Selective sweeps leave a tell-tale signature on the genome, with a reduction in genetic diversity at the locus under selection. Variants linked to the selected locus will hitchhike alongside the selected locus, increase in frequency and reduce genetic diversity (Kaplan et al. 1989; Smith & Haigh 1974). Reduction in genetic diversity diminishes with increasing genetic distance from the locus under suppression as recombination erodes the association between selected and unselected variants each generation (Braverman et al. 1995). Therefore, analysing the genetic impact of a selective sweep can help to inform us about the location of the selected locus; with the selected locus being identified by the greatest reduction of genetic diversity and change in frequency of the allele linked to the sweep for suppression (relative to other loci).

Analysis of the selective sweep for suppression of male-killing in *H. bolina* on Upolu revealed that the most intense signature of selection was located around the Hb99/Hb175/Hb176 cluster of markers that lies in the middle of the suppressor region. This indicates that the suppressor locus is located closer to these markers than it is to other examined markers.

This chapter seeks to identify a candidate gene for suppression. It does this firstly by creating new markers within the suppressor region. These markers are mapped and used for further analysis to narrow down the suppressor region. This is achieved through further segregation analysis, and then by further analysis of the selective sweep for the suppressor on Upolu. The suppressor region is deemed to be the area shown as necessary for male survival, and that contains the strongest signature of selection associated with the selective sweep for the suppressor.

This chapter then goes on to assess the hypothesis that *doublesex* is a candidate gene for the suppressor. That male-killing results from interference in the sex determination pathway

(Fukui et al. 2015; Sugimoto et al. 2015), leads to aberrant splicing of *doublesex* (Fukui et al. 2015; Sugimoto & Ishikawa 2012), and is found to be located orthologous to the suppressor region in other Lepidoptera (Hornett et al. 2014) does indeed make it an interesting candidate.This hypothesis is investigated by mapping *doublesex* in *H. bolina*, where it is shown to be located within the suppressor region. Intensity of selection at this locus is examined and changes to *doublesex* in suppressed and unsuppressed populations are then assessed.

Methods

Marker development and amplification

Markers were developed to further delineate the suppressor region and to investigate *doublesex* as a candidate gene. The *B. mori* genome was used to select markers located orthologous to the *H. bolina* suppressor region, as conservation of synteny is common amongst Lepidoptera (d'Alencon et al. 2010; Pringle et al. 2007). Coding sequence, derived from SilkDB (Duan et al. 2010; Wang et al. 2005), of the selected genes was BLASTed against the *H. bolina* NGS database (described in chapter 2). Contigs that were returned without paralogues were aligned to *B. mori* sequence to estimate intron/exon boundaries. Primers were designed to amplify the selected region with the aid of Oligo6.

PCR amplification of these markers was performed in a total volume of 12 µl, made up of 4µl ddH_2O, 6µl of 2x GoTaq Hot Start Green Master Mix (Promega), 0.5pmol forward primer, 0.5pmol reverse primer and 1 µl of DNA template. PCR cycling conditions consisted of an initial denaturation period of 1 minute at 95°C, followed by 35 cycles of 15 seconds at 95°C, 15 seconds at annealing temperature, then 72°C for the optimised extension time. This was followed by a final extension of 7 minutes at 72°C. Optimised annealing temperatures and extension times for each marker are specified in table 3.1.

PCR products were visualised on an agarose gel. Successfully amplified samples were then purified and sequenced using Sanger method following the PCR product and sequencing protocol (appendix ii). Sequence was aligned using Geneious Pro v.5.6.6 and scored for polymorphisms (SNPs, indels) that varied within and between populations.

Marker	F Primer (5'>3')	R Primer (5'>3')	F/R	Ta °C	Extension Time (sec)	*B. mori* Accession
Hb5100	GGAGCATGCCCTTAAGTAAGTTTC	CGGGTCCGTTCGTTATCTG	R	58	45	XM_012689849.1
Hb 5105	CCATCCGGAACCTCCTCTG	GTTGGCGAATCACTTTGTAACG	F	53	90	XM_012689568.1
Hb 5114	AATCGCTTTGTAGTTCAAGTGCTG	TGTTGAGTCCTTTCTGGATTTGG	F	60	75	XM_004921836.2
Hb doublesex	ATCATAATTGGAATATAGRCTTTYC	GCTTCCTGGCGTACTCGTTGATG	F	53	30	NM_001043406.1
Hb Prospero	GGATAGCGGACCCAGAAGAACAT	CCACCGACTTGCCTTACGA	F	57	60	XM_012689134.1

Table 3.1: Marker information: Primer information for markers developed and used in this study. Optimal annealing temperature (Ta) and extension time, which primer was used for sequencing and the *B. mori* accession number for orthologous gene used to design markers.

Enhancing linkage map of the suppressor region

Locations of the new markers within the *H. bolina* genome were mapped using the female daughters from the male-informative mapping families (a backcross design described in Hornett *et al*., 2014, and chapter 2). There are a total of 307 daughters in the male-informative mapping family. However, not all daughters are informative in terms of mapping marker locations, due to a lack of recombination events. To this end, each marker was mapped using a subset of daughters that had previously been identified as being recombinants (table 3.2).

Marker	n
Hb5100	90
Hb5105	28
Hb5114	80
Hbdoublesex	303
HbProspero	45

Table 3.2: Number of male informative family daughters used for mapping.

Mapping took place by amplifying and sequencing each marker using the male-informative family parents (mapping families are described in chapter 2 and Hornett *et al*., (2014)). To ascertain that markers were suitable for mapping purposes, the resulting sequence was assessed visually and SNPs that were present in male parents, but for which the female parent was homozygous, were used for mapping purposes. Markers were then amplified and sequenced, again via Sanger technology, in the female offspring of the male-informative family. Each daughter was scored for the paternal allele and the information used to create a new linkage map using Joinmap (Stam 1993). A graphical representation of this map was created using MapChart (Voorrips 2002).

Narrowing down the suppressor region through segregation analysis

The genomic region necessary for male survival was previously ascertained by looking at allele segregation in the 60 sons of the male-informative family (Hornett *et al*., 2014). A region can be identified as being necessary for male survival if a paternal allele is present in all of the sons. The occurrence of a recombination event in a son would demonstrate that the marker in question was not required for male survival and so that marker can be excluded from being located within the suppressor region. The male-informative family sons have previously been assessed for recombination events on chromosome 25 (Hornett *et al*., 2014). To further narrow down the suppressor region, 14 recombinant sons were sequenced with the newly-developed suppressor region markers. Allele segregation was scored and recombination events noted.

Narrowing down the suppressor region using selective sweep data

The *H. bolina* population on the island of Upolu, Independent Samoa had, as a result of *Wolbachia* infection, been extremely female-biased for at least 100 years up until 2001 (Dyson & Hurst 2004). The arrival of the suppressor into the population, sometime around 2005 led to a rapid selective sweep for the suppressor. By happenstance, a survey of *H. bolina* had been conducted, and samples collected in 2001, shortly before the arrival of the suppressor (Dyson & Hurst 2004). Samples were then collected from Samoa in 2006 just after the introduction of the suppressor and in 2010 sometime after the arrival of the suppressor. The genomic impact this selective sweep had on the population of Upolu, one of the two main islands that constitutes Independent Samoa, has previously been analysed. (Charlat et al. 2007b; Hornett et al. 2014).

Positive selection events leave a distinct signature on the genome, which can be used to help to localise the genomic region under selection. The selected allele will increase in frequency within the population, reducing genetic diversity. Alleles at other loci that are linked to the allele under selection will hitchhike along with the selected allele, reducing diversity at these loci (Kaplan et al. 1989; Smith & Haigh 1974). This signature of selection is strongest at the locus under selection but, as a result of recombination, diminishes in loci

more weakly linked to the selected locus and also diminishes over time (Braverman et al. 1995). The genomic region carrying the suppressor can thus be narrowed down by assessing which markers show the greatest signature of selection i.e. the greatest magnitude of change of allele frequency distributions. Markers distal to those showing the greatest signature of selection can be eliminated from belonging to the suppressor region.

The Upolu dataset was elaborated upon by sequencing the samples from 2001 (n=48), 2006 (n=48) and 2010 (n=46) with the 5 markers developed in this study. The resulting sequence was aligned using Geneious Pro v.5.6.6 (Kearse et al. 2012) and polymorphisms present within the populations were identified and scored for each individual. Haplotypes for each dataset were reconstructed using PHASE v2.1 (run with 1000 iterations, thinning interval of 100 and 1000 burn-in iterations) (Stephens & Scheet 2005; Stephens et al. 2001). Allele frequencies for each haplotype were then calculated for each population and alleles identified that had increased in frequency in the post-suppressor populations, compared with the pre-suppressor population. To assess the magnitude of change in genetic differentiation for each of the population pairs, pairwise F_{ST} between the 3 populations was calculated for each marker in Genepop v4.5.1 and significance calculated using the exact G test, again in Genepop v4.5.1 (Raymond & Rousset 1995; Rousset 2008).

Comparison of doublesex marker in suppressed vs unsuppressed wild populations

To assess *doublesex* as a candidate locus, the Hbdoublesex marker was amplified and sequenced in wild-caught samples from a number of populations in which the suppressor was present or absent (table 3.3). Each individual was first scored for any polymorphisms and haplotypes reconstructed using PHASE, as described above. A conspicuous indel marker with a 2bp AA insertion was identified. Each individual was scored as being one of three possible variants at this locus: homozygous long, homozygous short, or heterozygous. Data from wild populations in which the suppressor was absent was pooled and compared to pooled data for populations in which the suppressor was present.

Population	Suppressor Status	N	Sample Reference
Vietnam (2006)	Present	8	Charlat *et al*., (2005)
Borneo (2001)	Present	13	Charlat *et al*., (2005)
Upolu, Independent Samoa (2001)	Absent	41	Dyson & Hurst (2004); Hornett *et al*., 2014
Upolu, Independent Samoa (2006)	Present	41	Charlat *et al*., (2007); Hornett *et al*., 2014
Upolu, Independent Samoa (2010)	Present	46	Hornett *et al*., 2014
Savai'i, Independent Samoa (2001)	Absent	23	Charlat *et al*., (2005)
Savai'i, Independent Samoa (2007)	Present	15	Unpublished
Savai'i, Independent Samoa (2010)	Present	39	Unpublished

Table 3.3: Wild-caught samples used to investigate Hbdoublesex marker.

Results

Enhancing linkage map of the suppressor region on chromosome 25

An updated linkage map of *H. bolina* chromosome 25 that included the markers Hb5114, Hbdoublesex, HbProspero, Hb5105 and Hb5100 was created (figure 3.2). Mapping resolution did not separate HbProspero and Hb5105, which co-segregated. Note that in *Papilio* butterflies, the lepidopteran genus with an available genome that has the greatest level of gene order synteny with the *H. bolina* suppressor region, *prospero* is located closer to *doublesex* than it is to Hb5105 (Kunte et al. 2014; Nishikawa et al. 2015); it is expected that this may also be the case for *doublesex* in *H. bolina*.

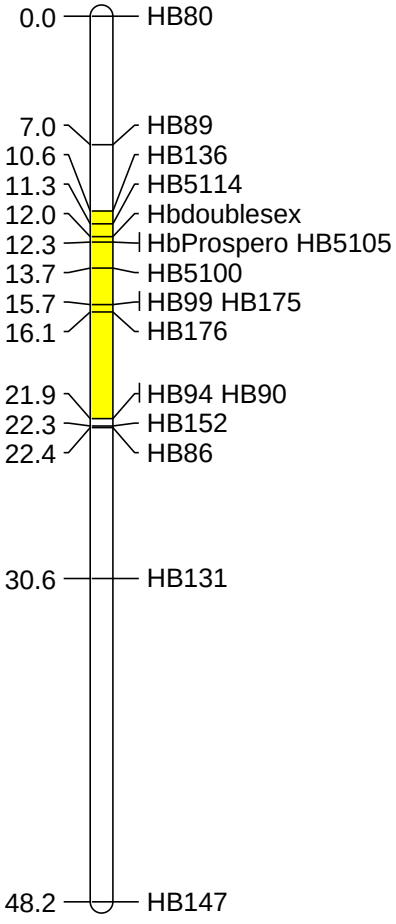

Figure 3.2: Linkage map of *H. bolina* chromosome 25. Graphical representation of *H. bolina* chromosome 25. Numbers represent distance from the distal marker (Hb80) in cM. The region containing the suppressor, as identified by Hornett *et al*. (2014), lies between Hb136 and Hb94/Hb90 and is highlighted in yellow.

Narrowing down the suppressor region using segregation analysis

The suppressor region can be identified via segregation analysis using the male-informative family – the presence of a paternal allele in all sons would deem that region as being necessary for male survival. The recombination event in the male informative son 5100 was shown to extend to marker Hb5114. This demonstrates that the 0.7cM region that lies between Hb136 and Hb5114 is not required to suppress male-killing. The paternal allele was present in all other individuals tested for the five markers i.e. it co-segregated with suppression, and so proved ineffective at further narrowing down the opposite end of the suppressor region.

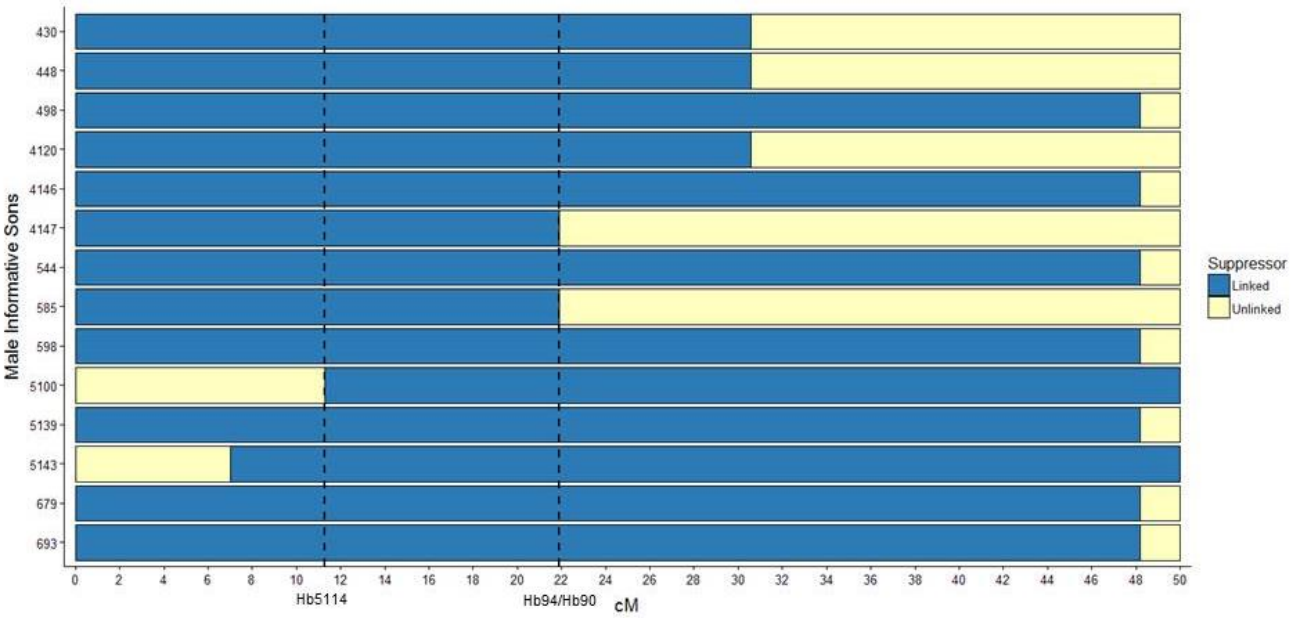

Figure 3.3: Recombinational breakpoints in male informative family sons along chromosome 25. Sons from the male-informative family that were found to contain recombination events along chromosome 25 are represented on the y-axis. The x-axis represents *H. bolina* chromosome 25, with positions shown in cM. Blue blocks represent regions linked with male survival, light-coloured blocks represents genetic regions not necessary for male survival. Recombinational breakpoints are indicated by a change in block colour. The genetic region associated with suppression is marked by vertical dashed lines, and lies between marker Hb5114 (11.3cM) and Hb90/Hb94 (21.9cM).

Narrowing down the suppressor region using selective sweep data

Wild *H. bolina* samples were collected from Upolu, Independent Samoa at three time points before (2001) and after (2006 and 2010) a selective sweep for the suppressor. Theory predicts that markers most tightly linked to the suppressor locus should show the greatest signature of selection.

The greatest change in allele frequencies was found to have occurred at the HbProspero marker, some 3.5cM away from the previously identified peak in the Hb99/Hb175/Hb176 cluster. At HbProspero, a novel suppressor-linked allele appeared in the 2006 population at 71% frequency (figure 3.5). Furthermore, there was an ongoing increase in the suppressor-linked allele between 2006 and 2010 by which time this allele had risen to 95% frequency (figure 3.5). This marker also showed the greatest change in genetic diversity between pre-sweep and post-sweep populations, as measured by F_{ST} (table 3.4).

Marker	2001 x 2006	2001 x 2010	2006 x 2010	Reference
HB80	0.054 ***	0.049 ***	-0.007	Hornett *et al.*, 2014
HB89	0.170 ***	0.163 ***	0.014	Hornett *et al.*, 2014
HB136	0.097 **	-0.007	0.075 **	Hornett *et al.*, 2014
Hb5114	0.223 ***	0.303 ***	0.044 **	This study
Hbdoublesex	0.183 ***	0.206 ***	0.000	This study
HbProspero	0.318 ***	0.560 ***	0.128 ***	This study
Hb5105	0.295 ***	0.317 ***	0.013 **	This study
Hb5100	0.135 ***	0.123 ***	0.029 *	This study
HB99	0.204 ***	0.180 ***	-0.005	Hornett *et al.*, 2014
HB175	0.190 ***	0.152 ***	0.001	Hornett *et al.*, 2014
HB176	0.286 ***	0.315 ***	0.000	Hornett *et al.*, 2014
HB90	0.116 ***	0.023 *	0.040 **	Hornett *et al.*, 2014
Hb94	0.074 ***	0.084 ***	0.060 ***	Hornett *et al.*, 2014
HB152	0.149 ***	0.045 ***	0.026	Hornett *et al.*, 2014
Hb86	0.105 ***	0.036 **	0.032 *	Hornett *et al.*, 2014
HB131	0.116 ***	0.135 ***	-0.009	Hornett *et al.*, 2014
HB147	-0.001	0.014 ***	-0.006	Hornett *et al.*, 2014

Table 3.4: F_{ST} values between population samples from different timepoints at each marker along chromosome 25. Standardised genetic differentiation between populations from 3 time points: 2001, 2006, 2010. Deviations that are statistically significant from F_{ST}=0 are denoted by *** ($p<0.001$), ** ($p<0.01$), * ($p<0.05$).

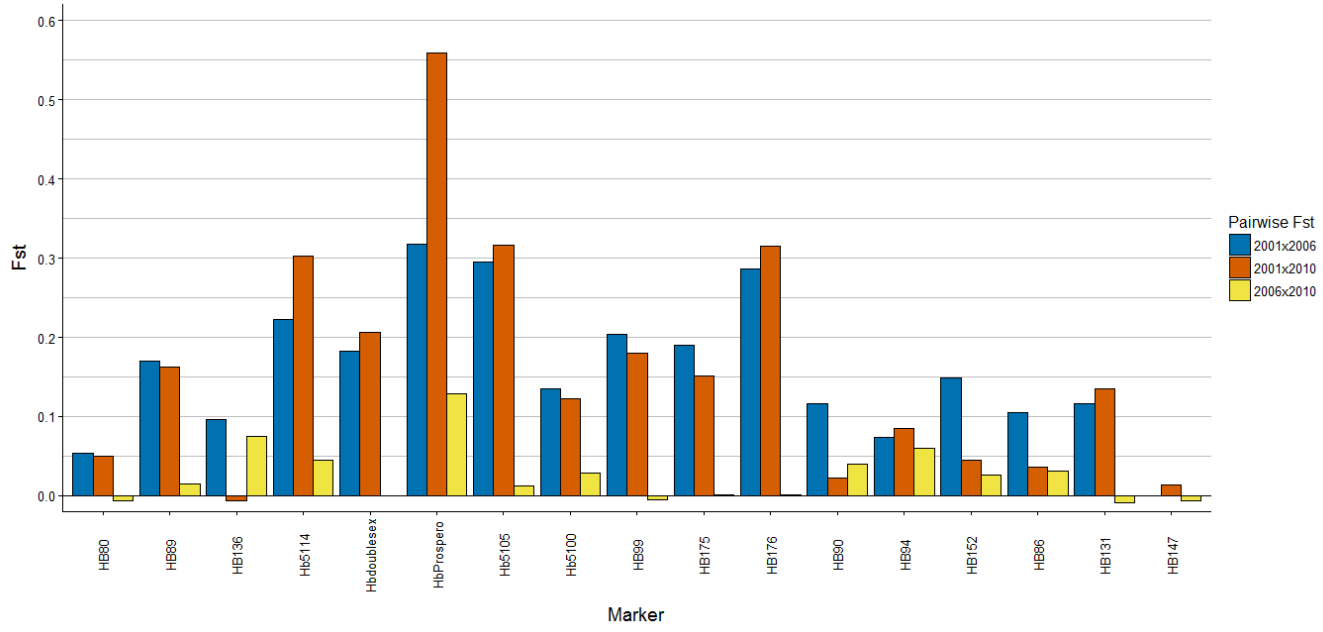

Figure 3.4: F_{ST} standardised population genetic differentiation at markers on chromosome 25 for populations from Upolu, Independent Samoa at 3 different time points. 2001, 2006, and 2010.

After HbProspero, the Hb5105 marker had the most intense signature of selection. The primary suppressor-linked allele at this marker is novel in post-sweep populations, where it rises to 58% frequency. Changes in genetic diversity between pre-sweep and post-sweep populations, as measured by F_{ST} (table 3.4) were also less profound than at the HBProspero locus. This difference occurs in spite of the fact that the HbProspero and Hb5105 marker co-segregate in the mapping family.

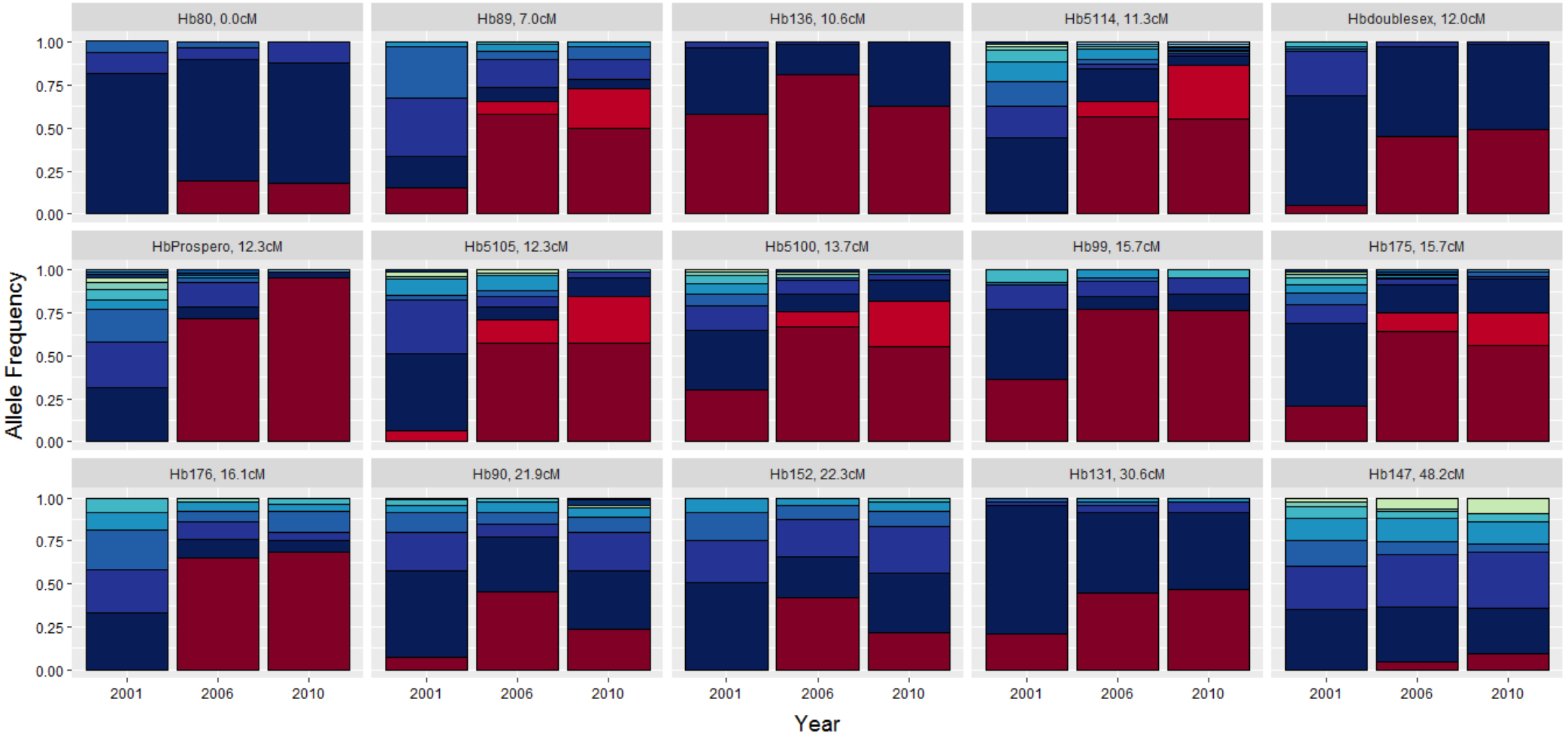

Figure 3.5: Change in allele frequencies across *H. bolina* chromosome 25. Allele frequencies of *H. bolina* collected from Upolu, Independent Samoa in 2001, 2006 and 2010 at 15 markers along *H. bolina* chromosome 25. Allele frequency is shown on the y-axis and year of sample (2001, 2006, 2010) along the x-axis. The major allele that increased in frequency with the introduction of the suppressor is coloured in red, secondary alleles that also increased in frequency are coloured in a lighter shade of red. Haplotypes that did not increase in frequency are shown in blue. Markers are ordered relative to their position along the chromosome, with position indicated in cM in the title of each plot.

The strongest signature of selection, associated with suppression of male-killing, is found at HbProspero and Hb5105 but then diminishes at more distal markers. The one exception to this is for the Hb176 marker, where changes in genetic differentiation were higher than at the more closely-linked Hb175 and Hb99 markers but still less intense than at HbProspero. This indicates that the locus of suppression is most likely located closer to the HbProspero and Hb5105 markers than it is to the Hb5100, Hb99 and Hb175 markers.

When evidence from recombinational analysis is taken along with analysis of the selective sweep for the suppressor, it suggests that the genetic region lying between markers Hb5114 and Hb5100 is the region in which the suppressor mutation is most likely to reside (figure 3.6). However, as the exact genetic change that controls suppression has not yet been identified it is not possible to say for certain that the suppressor does reside within this region.

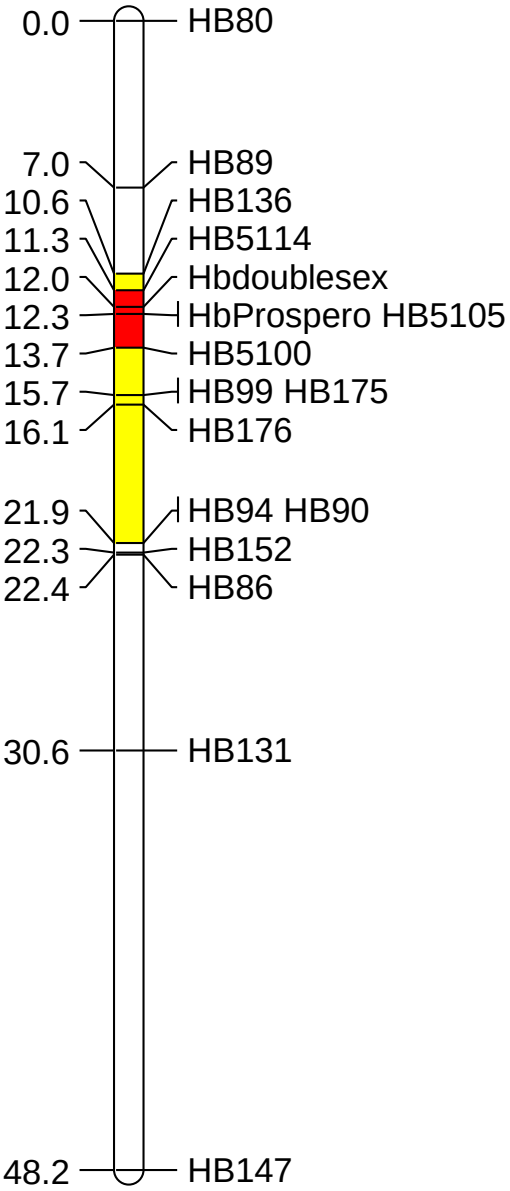

Figure 3.6: Linkage map of *H. bolina* chromosome 25 showing suppressor region. Graphical representation of *H. bolina* chromosome 25. Numbers represent distance from the distal marker (Hb80) in cM. The region containing the suppressor lies between markers Hb5114 and Hb5100 and is highlighted in red. The previous suppressor region, as identified by Hornett *et al*., (2014), lies between Hb136 and Hb94/Hb90 and is highlighted in yellow.

doublesex as a candidate for suppression

The *doublesex* marker was mapped and found to be located within the suppressor region on *H. bolina* chromosome 25. Upon sequencing in the male-informative family, a paternal allele

was found to be present in all sons: *doublesex* co-segregates with suppression of *Wolbachia*-induced male-killing.

Even though the *doublesex* locus is positioned 0.3cM away from HbProspero, the marker with the greatest magnitude of change in allele frequencies, the selective sweep for the suppressor on Upolu seemingly had less of an impact at the *doublesex* locus. Pairwise F_{ST} value between pre-sweep population and post sweep populations (2001 vs 2006: F_{ST} =0.183, 2001 vs 2010: F_{ST} =0.206) are appreciably lower than at the neighbouring markers: Hb5114, HbProspero, Hb5105. Suppressor-linked alleles also didn't increase in frequency as much as at neighbouring markers. Closer examination of changes in allele frequencies at *doublesex* reveals that one allele present in the 2001 population sample decreases in frequency from 64% in 2001 to 53% in 2006 and 50% in 2010. Another allele is shown to increase in frequency from 5% in 2001 to 45% in 2006 and 49% in 2010. These two alleles are in equilibrium with each other.

This pattern is replicated when pooling data from wild caught samples from suppressed (n=162) and unsuppressed (n=64) populations (table 3.5) for the Hbdoublesex marker. In suppressed populations 2 alleles reach a balanced equilibrium; higher levels of allele diversity are found in wild, unsuppressed populations (figure 3.7).

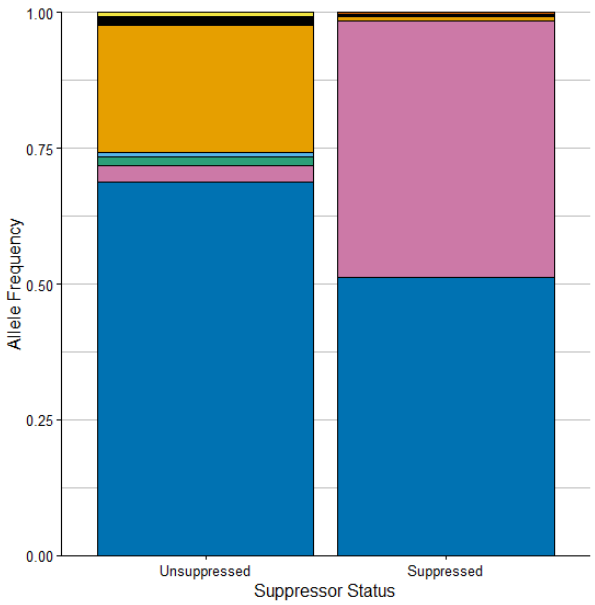

Figure 3.7: Frequency of allele diversity at Hbdoublesex marker in wild *H. bolina* populations in which the suppressor is either present or absent.

Sequence traces of the Hbdoublesex marker arising from wild-caught samples was scrutinised. A conspicuous change could be seen in the samples from post-suppressor populations versus those from pre-suppressor populations. This change was the presence of a conspicuous indel, the result of a 2bp AA insertion event. All wild-caught samples were scored as being either homozygous short (2 copies of 5A sequence), homozygous long (2 copies of 7A sequence) or heterozygous short/long (1 copy each of 5A allele and 7A allele). Whilst the heterozygous genotype was only found in 4 out of 64 individuals from the pre-suppressor populations, it was present in 155 out of 162 wild-caught individuals from post-suppressor populations (table 3.5, figure 3.8).

Population	Suppressor Status	SS	SL	LL
Vietnam	Present	0	8	0
Borneo	Present	0	11	2
Upolu, 2001	Absent	37	4	0
Upolu, 2006	Present	3	38	0
Upolu, 2010	Present	0	46	0
Savai'i, 2001	Absent	23	0	0
Savai'i, 2007	Present	0	15	0
Savai'i, 2010	Present	2	37	0

Table 3.5: Genotypes in wild *H. bolina* populations at Hbdoublesex marker. Wild-caught *H. bolina* from populations in which the suppressor was either present or absent. Individuals were scored as being either homozygous short (SS), heterozygous (SL), or homozygous long (LL) for an indel in the Hbdoublesex marker.

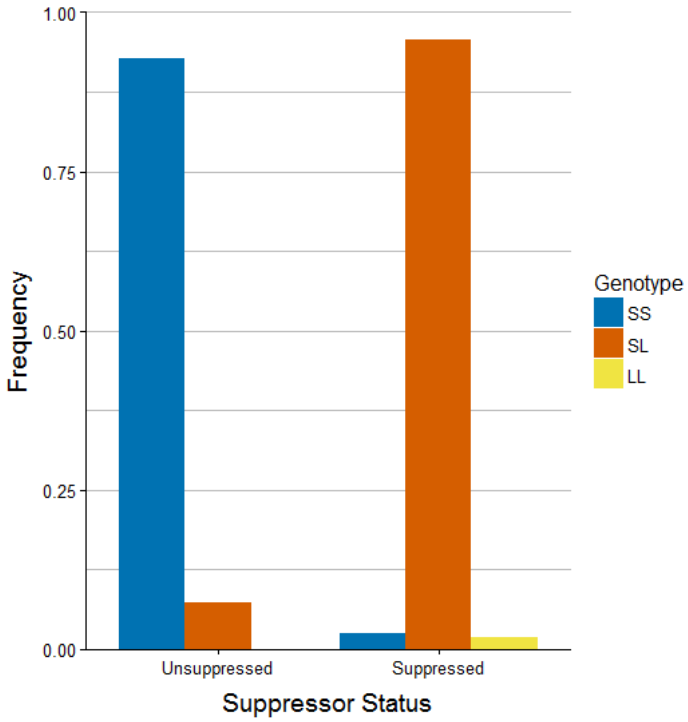

Figure 3.8: Genotype frequencies for a conspicuous indel marker in wild *H. bolina* populations. Individuals in populations in which the suppressor is either present or absent were scored as either being homozygous long, homozygous short or heterozygous for a 2bp AA insertion. Unsuppressed populations n=64, suppressed populations n=162.

To investigate whether this excess of apparent heterozygotes in wild populations was due to the marker not being able to segregate in a Mendelian fashion, male and female offspring from the male-informative mapping family were sequenced at this marker (table 3.6). If the marker is Mendelian then there should be a 1:1 ratio of individuals homozygous for the short allele to individuals heterozygous for the allele in the male-informative family daughters.

	SS	SL	LL
Male	1	39	0
Female	121	178	0

Table 3.6: Genotypes at Hbdoublesex marker in male-informative mapping families. Sons and daughters from the male-informative mapping families were scored as being either homozygous short (SS), heterozygous (SL), or homozygous long (LL) for an indel in the Hbdoublesex marker.

Results showed that the Hbdoublesex marker does segregate in female offspring, although not quite in a Mendelian fashion, with there being somewhat more heterozygotes than expected, at a ratio of 1.5:1 heterozygotes to homozygotes. All bar one male offspring was found to be heterozygous for the indel.

Discussion

The blue moon butterfly *Hypolimnas bolina* has evolved the ability to suppress *Wolbachia*-induced male-killing, so that infected females are able to produce viable male and female offspring. An approximately 10cM region on *H. bolina* chromosome 25 has previously been identified as bearing the mutation that permits suppression to occur (Hornett *et al*., 2014). This chapter aimed to more precisely identify the location of the suppressor region and to investigate the hypothesis that the *doublesex* gene is a candidate locus of suppression. The outcome of this investigation can be summarised thus:

Narrowing down the suppressor region:

- An updated linkage map of *H. bolina* chromosome 25 was created that included an additional five markers within the suppressor region.
- Recombinational analysis eliminated the 0.7cM region between markers Hb136 and Hb5114 from the suppressor region.
- The most intense signature of selection in the Samoan selective sweep vent was found at the HbProspero/Hb5105 markers with intensity of selection diminishing at more distal markers.

  This indicates that the 2.4cM region between markers Hb5114 and Hb5100 is the region in which the suppressor mutation is most likely to reside.

doublesex as a candidate for suppression:

- The *doublesex* gene was found to co-segregate with suppression, and lies within the region identified above through analysis of a selective sweep
- *doublesex* isn't the locus with the most intense signature of selection, in response to a selective sweep for the suppressor.
- There is an excess of individuals that are apparently heterozygous at *doublesex* in wild, suppressed populations.

Narrowing down the suppressor region

The region in which the genetic change that controls the suppression of male-killing is located was identified as most probably being present within the 2.4cM region of *H. bolina* chromosome 25 that lies between the Hb5114 and Hb5100 markers. Gene order synteny in

the Asian swallowtail *Papilio xuthus*, in the region orthologous to the suppressor region is more similar to that found in *H. bolina* than in any other lepidopteran species, for which assembled whole genome data is available. In *P. xuthus* the region orthologous to the suppressor region is 6.4Mbp in length and contains 23 genes, inclusive of the Hb5114 and Hb5100 orthologues. The equivalent region in *B. mori* is 1.2Mbp and contains 27 genes.

The newly-defined suppressor region was characterised using five markers, of which HbProspero underwent the greatest magnitude of change between pre-suppressor and post-suppressor samples. In the *H. bolina* chromosome 25 linkage map, HbProspero co-segregates with Hb5105. In other Lepidoptera this marker has been shown to lie closer to Hb5100 and owing to high levels of gene order synteny found across the Lepidoptera, it is likely that that this may also be the case with these two markers in *H. bolina*. This coupled with a greater intensity of selection at HbProspero than at Hb5105, suggests that HbProspero is the marker closest to the locus of selection.

Furthermore, the suppressor-linked allele at HbProspero showed ongoing selection between 2006 and 2010 to a much greater extent than at any other marker, where the frequency of the allele associated with the introduction of the suppressor had either remained stable or decreased between 2006 and 2010. When male-killing is suppressed, a CI phenotype is uncovered which drives *Wolbachia* towards fixation (Hornett et al. 2010). This means that selection pressure for the suppressor remains and models predict that the suppressor should fix in the population (Hornett *et al*., 2014). The observation that selection for the suppressor-linked allele at HbProspero was ongoing, even after the population sex ratio of Independent Samoa had returned to parity, shows that this locus must be closely linked to the locus under selection.

Interestingly, there are a number of markers (Hb5105, Hb5114, Hb5100, Hb89, and Hb175) where a secondary allele, often novel, also increased in frequency with the introduction of the suppressor. All of these secondary alleles were seen to increase in frequency between 2006 and 2010. This may show that hitchhiking is still occurring at these markers but that recombination events since the appearance of the suppressor onto Upolu had altered the allele associated with the locus under selection.

The appearance of novel alleles associated with suppression appearing in Samoan populations (coupled with shared genotypes at *doublesex* in suppressed individuals from both South-East Asian and Polynesian populations) indicates that the suppressor is spreading through the species range via migration, rather than recurrent *de novo* mutation.

doublesex as the target of selection

In this chapter the hypothesis that the *doublesex* gene may be the target of selection for suppression of *Wolbachia*-induced male-killing in *H.bolina* was investigated. It was first established that as hypothesised, *doublesex* was indeed located within the suppressor region and also co-segregated with male survival.

It was hypothesised that the evolution of suppression resulted from changes to the *doublesex* gene. Although the *doublesex* marker is located in the region of chromosome 25 that shows the greatest signature of selection, it was not the marker that showed the greatest signature of selection in response to the selective sweep for the suppressor on Samoa, which would have been expected if *doublesex* was the target of selection. Rather, *doublesex* showed a less intense response to selection than at surrounding markers. At *doublesex*, two alleles were found to reach a balanced equilibrium in both the 2006 and 2010 time points, which resulted from almost every individual in post-suppressor populations from having the same genotype. This phenomenon was also found in wild-caught samples from South-East Asian populations (Vietnam and Borneo), even though the suppressor has been established in these populations for much longer than in Independent Samoa.

All males except one individual were heterozygous at the *doublesex* marker. This marker is an indel that was deemed to be an AA insertion event located a few hundred basepairs upstream of exon three in the *doublesex* gene. It is possible that this indel could result in a frameshift mutation that would alter the expression of *doublesex*. However, there is no evidence of nonsynonymous changes within doublesex associated wither with suppression or the presence of this indel, suggesting that the indel had any functional consequences on *doublesex*.

If *doublesex* is the locus of selection then all surviving males would be expected to carry the *doublesex* indel marker. However, this is not the case as not all surviving males possess this marker. The high proportion of surviving males possessing this marker does however suggest that it must be closely linked to the locus necessary for male survival. This could mean that whilst the marker is closely linked to the locus of selection, it is not in itself the genetic change responsible for male survival. Furthermore, male-killing is not fully penetrant and about 1 in 100 male offspring manage to avoid male-killing even though their mother was infected with a male-killing strain of *Wolbachia*. If this individual is an 'escapee' then it would be able to reach maturity without possessing the suppressor allele.

In wild, suppressed populations individuals heterozygous for the *doublesex* indel existed at a relatively low frequency (0.17) in populations in which the suppressor was absent, heterozygous individuals were found at a high frequency (0.96) in populations in which the suppressor is present. Not only is this a remarkable change in genotype frequencies, but it is also unexpected. If these heterozygous individuals were mating without restriction in the wild then individuals homozygous for each of the allele types would be expected to occur, but they don't. That individuals homozygous for each *doublesex* allele do occur in wild populations and that the locus does segregate into homozygotes and heterozygotes, albeit with an excess of heterozygotes, in the mapping family suggests that the excess of heterozygotes does not result from the locus being homozygous lethal in suppressed individuals. Why then are there so many heterozygotes in wild populations?

The excess of heterozygotes at *doublesex* in suppressor present populations could be explained by a duplication event associated with suppression of male-killing. If *doublesex*, or part of the *doublesex* gene was duplicated so that each individual from a population with the suppressor now contained two copies of *doublesex* (on each chromosome) and that each of these copies was different then when the Hbdoublesex marker was amplified and Sanger sequenced it would appear that every individual were heterozygotes, as they had 2 different alleles, when in fact they were all homozygous for 2 different variants of *doublesex* (figure 3.9). This excess of 'apparent heterozygotes' has been observed to occur in a selective sweep for gene duplication that promotes insecticide resistance in *Culex pipiens* (Lenormand et al. 1998).

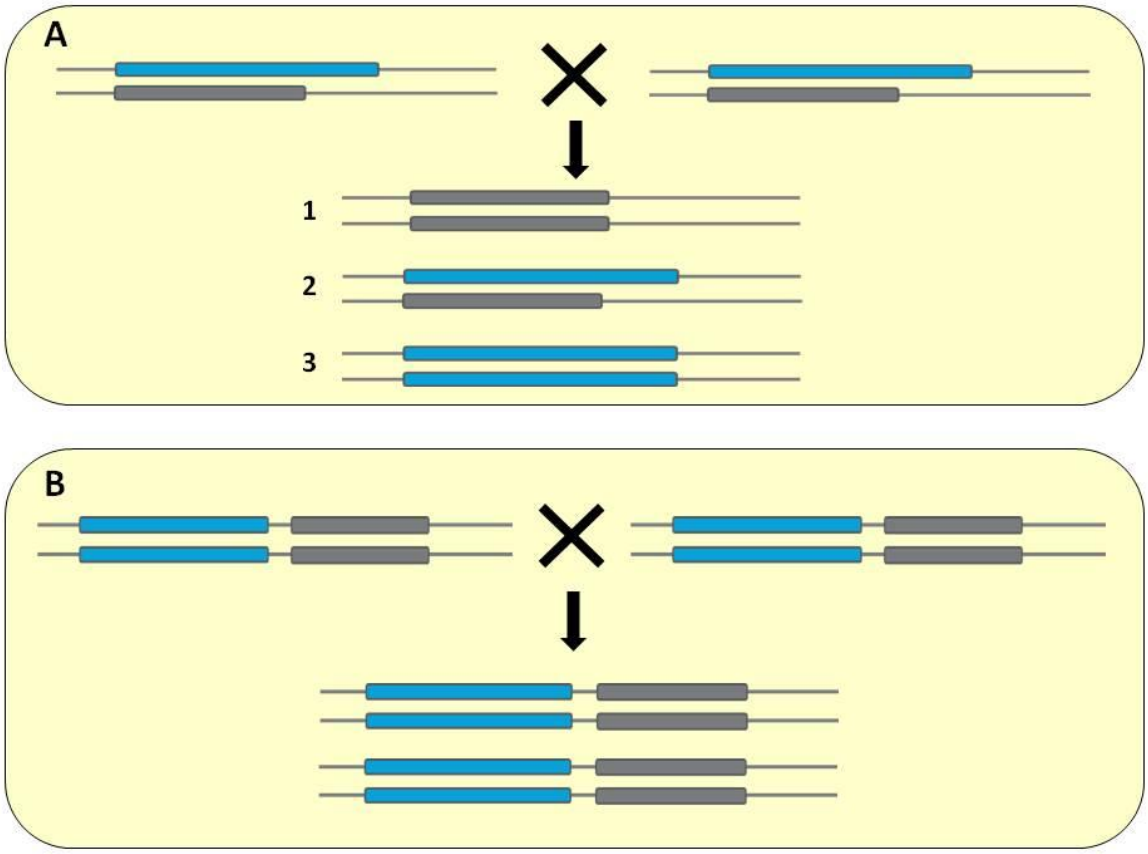

Figure 3.9: excess of heterozygotes in wild samples could be explained by duplication of *doublesex*. Panels A and B represent a cross between parents that have 2 different alleles at a locus and the offspring that should be expected from such crosses. Coloured blocks represent alleles. Panel A illustrates a cross between parents that are heterozygous, having 1 copy of 2 different alleles at the locus. Such a cross would be expected to produce 3 types of offspring at this locus: 1. Homozygous for the short allele; 2. Heterozygous with 1 copy of the short allele and 1 copy of the long allele; homozygous for the long allele. However, types 1 and 3 are very rarely found to occur in wild populations of *H. bolina* in which the suppressor is present.
Panel B illustrates a cross in which each parent is homozygous, having 2 copies of 2 different alleles. Only 1 type of offspring would be expected to result from such a cross: individuals that are also homozygous for 2 different alleles at this locus. Hence duplication of *doublesex* could explain the excess of individuals that are apparently heterozygous at *doublesex* in wild, suppressed populations.

If suppression of male-killing is indeed associated with a duplication of *doublesex* then re-examining the selective sweep haplotype frequency information tells a somewhat different story. Now, rather than two different alleles reaching a balanced equilibrium these alleles actually constitute a single suppressor-associated haplotype. This haplotype attains a frequency of 97.5% in the 2006 Upolu population, rising to 99% frequency in the 2010 population, which is higher than the suppressor-linked allele at any of the other examined markers. Models of this system have previously predicted that the locus of selection would reach fixation.

Furthermore, if a duplication involving *doublesex* has taken place then the nature of this event is more likely to be a tandem duplication than it is a duplication and translocation event. Evidence for this comes from that fact that the *doublesex* marker used to infer the occurrence of a duplication event was also used to map the location of *doublesex*. This marker mapped to a genomic region consistent with *doublesex* orthologues in other Lepidoptera. Segregation of this marker in the F_2 offspring male-informative families was also concurrent with *doublesex* mapping to a single genomic region i.e. the segregation pattern was consistent with other markers in the region and showed no unexpected or unusual patterns of segregation. As demonstrated in chapter two, only a single genomic region on chromosome 25 has been shown to be associated with suppression. If a *doublesex* paralogue has arisen elsewhere within the genome of suppressor-positive butterflies then it doesn't seem to be having a functional effect, or at least does not appear to be the primary determiner of male survival. There is no evidence to indicate that the 1.5:1 ratio of heterozygous to homozygous daughters found in the male-informative mapping family is the result of *doublesex* paralogues elsewhere in the genome.

Gene duplication is a commonly seen method of generating evolutionary novelty (Zhang 2003). So it makes sense as a mechanism through which suppression could have evolved. This result leads to hypothesis that suppression of *Wolbachia*-induced male-killing in *H. bolina* is associated with changes to *doublesex* that may include duplication of all or part of the gene.

Duplication of *doublesex* has been observed in crustaceans and chelicerates (Price et al. 2015). In these cases sex-specific alternative splicing doesn't occur rather there is sex-biased

expression in males. Whole or partial gene duplication of *doublesex* has also been found to have occurred in a number of insect families: the Zoraptera, Hymenoptera (*Sphaeropthalma orestes)*, and in a number of Coleoptera (*Meloe violaceus, Gyrinus marinus, Meligethes aeneus, Onthophagus nigriventris*) and Diptera (*Belgica antarctica*). This sets a precedent that shows that *doublesex* can and does duplicate and so is a possible hypothesis for evolutionary change in *H. bolina*. Gene duplication can lead to changes in expression level and presents the opportunity for evolutionary novelty. In the case of *H. bolina,* the former hypothesis is preferred. Male-killing in other Lepidoptera has seen to disrupt *doublesex* expression in male-killed individuals (Fukui et al. 2015; Sugimoto & Ishikawa 2012; Sugimoto et al. 2015). Upregulation of *doublesex* protein may alleviate that.

Future work

Future work should aim to establish the exact molecular changes that occurred to give rise to the evolution of suppression of *Wolbachia*-induced male-killing in *H. bolina*. The first step needed to achieve this would be to create an assembly of the *H. bolina* genome and to then create high quality assemblies of the suppressor region for an individual in which the suppressor is present and also for an individual in which the suppressor is absent.

A high-quality assembly of the suppressor region would enable new markers to be designed within the region. These could be used in conjunction with much larger male-informative mapping families, with hundreds of sons, to delimit the suppressor region through recombinational analysis.

The suppressor region assembly could also act as a template to delimit the suppressor through population genetic analysis. Sequencing pooled samples from populations in which the suppressor is present and from populations in which the suppressor is absent would permit direct comparison of the region. From this, the presence of any structural changes such as gene duplication or inversion and any sequence changes associated with suppression could be ascertained.

Further analysis of the Samoan selective sweep could also help to identify the genetic changes involved in the evolution of suppression. *Wolbachia* becomes fixed in populations

with the suppressor (Hornett et al. 2010), so the suppressor would still be under selection to enable male survival. Over time hitchhiking at loci distal to the locus of selection would have been broken down by recombination, so the signature of selection would be seen in a much narrower area around the suppressor. Collecting new *H. bolina* specimens from Independent Samoa and comparing them to the pre-suppressor Samoan samples across the suppressor region would pinpoint a narrower genetic area under selection.

Transcriptomic analysis of *doublesex*, and any other candidate genes identified from the proposed future scheme of work, should also be assessed to determine if there are any changes associated with suppression. To see what if any changes result from suppression, transcriptomic analysis should take place using male and females without *Wolbachia* or the suppressor, and in *Wolbachia*-infected male and females with and without the suppressor. If suppressor-associated changes are observed in a candidate gene then further analysis of expression levels at different stages of development would further understanding of the mechanisms through which the suppressor operates.

Conclusion

In this chapter, the genomic region in which the mutation responsible for the suppression of *Wolbachia*-induced male-killing in *H. bolina* was narrowed down to a 2.4cM region that encompasses the gene *doublesex*. This is the basal gene in the insect sex determination cascade and signals for cells to develop sexually dimorphic characteristics. This chapter proposes the hypothesis that duplication of *doublesex*, or of part of the *doublesex* gene, is associated with suppression of male-killing. If this is the case then it demonstrates that endosymbionts can act as drivers of evolutionary change in key developmental pathways, such as the sex determination pathway.

Chapter 4

Hitchhiking in space under extreme selection pressure

Abstract

Hypolimnas bolina butterflies on Independent Samoa are infected with a male-killing strain of *Wolbachia*. This led to an extremely female-biased population sex ratio of 100:1, which persisted for over a century. In 2005 however, male butterflies were seen more frequently on one of the islands, Upolu, and this was found to be associated with the spread of a mutation that rescued males (see chapter 3). Whilst the neighbouring island of Savai'i remained at a 100 female per male sex ratio in 2005, the sex ratio skew had begun to diminish by 2006, particularly in collections close to Upolu. The change in Savai'i is hypothesized as being associated with the wave of advance of the suppressor, which having spread through Upolu then arrived in Savai'i (10 miles away) by dispersal. This chapter analyses the selective sweep for the suppressor as it progressed in space, moving from the island of Upolu and onto Savai'i. This analysis revealed a very intense selection event. The breadth and intensity of the selective sweep was weaker on Savai'i than on Upolu, consistent with the timing of sex ratio changes. Both the Upolu and Savai'i sweeps were found to be centred around the *doublesex* gene and are consistent with the hypothesis proposed in chapter 3, that the evolution of suppression is associated with duplication of *doublesex*.

Data for the markers Hb80, Hb89, Hb99, Hb131, Hb152, and Hb175 for samples from Upolu is derived from Hornett *et al*., (2014) but is shown in this chapter in order to permit comparison to data pertaining to samples from Savai'i.

Introduction

When an advantageous mutation arises, selection will act to increase the frequency of the variant within a population. If positive selection for the new variant is sufficiently strong, then it will continue to increase in frequency towards fixation. During periods of positive selection, alleles at other loci that are linked to the variant under selection will also increase in frequency in a process termed 'genetic hitchhiking' (Smith & Haigh 1974). Hitchhiking causes a reduction in variation around the locus under selection (Kaplan et al. 1989; Smith & Haigh 1974), a process commonly termed a 'selective sweep'. The selective sweep process is also associated with generation of linkage disequilibrium (LD), as the originally selected haplotype increases in frequency (Kim & Nielsen 2004; Kim & Stephan 2002; McVean 2007; Przeworski 2002; Stephan et al. 2006). Over time LD between the locus under selection and unlinked loci is broken down by recombination. The closer a hitchhiked variant is to the locus under selection then the less opportunity there is for recombination. Over time the genomic region affected by the sweep narrows so that reduced diversity only persists in the region closely linked to the selected locus (Braverman et al. 1995).

In a structured population the time it takes for a favourable allele to become fixed is increased, relative to a panmictic population, as it takes longer for the variant under selection to move through the population, and this is expected to decrease the efficacy of hitch-hiking (Barton 2000; Slatkin & Wiehe 1998). A selective sweep moves through a population in an expanding wave (Fisher 1937). In a structured population, a beneficial allele needs to reach high frequency levels before the selective sweep expands into another deme via migration. This increases the amount of time it takes for a sweep to reach fixation, providing more time for recombination to break down hitchhiking, between the beneficial allele and linked variants, thus reducing the breadth and intensity of the sweep as it travels across space (Barton 2000; Kim & Maruki 2011).

Individuals at the tip of the wave front will largely dictate which alleles are involved in hitchhiking. They introduce an advantageous allele into a new region, where it undergoes positive selection. Alleles that were linked to the favourable allele within these individuals will then hitchhike alongside it and also increase in frequency within the region. Variants from behind the wave front coalesce with a small number of individuals that were once at

the front of the wave. This increases genetic drift, which slows the advance of the sweep (Brunet et al. 2006; Mueller et al. 2011) and permits neutral or deleterious alleles, that were originally at the tip of the wave, to increase in frequency (Excoffier et al. 2009; Klopfstein et al. 2006). It also affects genetic differentiation within the population.

A selective sweep could cause genetic differentiation within a structured population to either increase or decrease, depending upon circumstances. As association between selected and neutral alleles is broken down, sequential fixation of the selected allele as it moves across space could result in different alleles hitchhiking in different subpopulations. This would increase local genetic differentiation around the selected allele in populations that originally had low levels of differentiation (Bierne 2010; Kim 2013; Slatkin & Wiehe 1998). Contrarily, if subpopulations were originally highly differentiated then the loss of diversity around the selected locus would decrease differentiation (Bierne 2010; Faure et al. 2008; Santiago & Caballero 2005; Slatkin & Wiehe 1998).

Areas of a genome that have been subject to selection can be identified by searching for the signature of selective sweeps, namely a reduction in diversity and the existence of LD (Fay & Wu 2000; Kim & Stephan 2002; Nielsen 2005; Sabeti et al. 2002; Tajima 1989; Voight et al. 2006). However, our understanding of the signature of selection in space is primarily derived from models (Barton et al. 2013; Excoffier et al. 2009; Kim 2013; Kim & Maruki 2011; Santiago & Caballero 2005; Slatkin & Wiehe 1998). This is largely through necessity as, although the signature of some selective sweeps across space in wild populations has been examined (Bierne 2010; Faure et al. 2008; Hoekstra et al. 2004; Koehn et al. 1980; Lenormand et al. 1998), it is difficult to obtain the genetic data required to study the origin and progression of such sweeps. A selective sweep for a mutation that enables the butterfly *Hypolimnas bolina* to suppress the male-killing effects of *Wolbachia* is a case where this genetic data is available.

Hypolimnas bolina has evolved the ability to suppress *Wolbachia*-induced male-killing (Hornett et al. 2006). Suppression first arose in South-East Asian populations. Evidence shows that suppression was apparent in late 19th century Filipino populations but absent across other South-East Asian (Borneo) and Pacific populations, which exhibited female-biased sex ratios (Hornett et al. 2009). Suppression of the *Wolbachia* male-killing phenotype

is now ubiquitous across South-East Asian (Charlat et al. 2005) and Japanese (Mitsuhashi et al. 2004) populations.

More recently, the suppressor spread to Independent Samoa. The prevalence of *Wolbachia* infection on Independent Samoa has historically been very high, with 99% of females being infected. This led to a female-biased sex ratio of 100:1, which persisted for over 100 years (Dyson & Hurst 2004). Independent Samoa is made up of two main islands: Upolu and Savai'i. The suppressor appeared in *H. bolina* on the island of Upolu at some point between 2001 and 2005 (Charlat et al. 2007b). Selection for the suppressor was incredibly strong. A selective sweep for the suppressor occurred and rapidly spread first across Upolu and then across the neighbouring island of Savai'i. In 2006 the population sex ratio was 1:1 on Upolu and the part of Savai'i nearest to Upolu, but remained female biased in populations distant from Upolu. By 2007 the suppressor had also spread across Savai'i, and the sex ratio had returned to parity. On Savai'i we can time the episode of selection more precisely than on Upolu, with the the suppressor mutation causing the population sex ratio to change from 100:1 (2005) to 1:1 (2006-7) in fewer than 10 generations (Charlat et al. 2007b).

Hypolimnas bolina samples were collected from both Upolu and Savai'i in 2001, prior to the selective event, in 2006 immediately after the sweep on Upolu and during the sweep on Savai'i, in 2007 once the sweep had progressed across both Upolu and Savai'i and then again in 2010, 4 years or around 40 generations after the selective sweep. These samples, uniquely, permit analysis of a selective sweep during its initial stages and also of how a selective sweep occurs over space.

The genetic signature left by the sweep for the suppressor as it progressed across Upolu, was examined by Hornett *et al*., (2014). Intense selection for the suppressor was accompanied by an equally intense signature of selection at the genetic level. There were strong hitchhiking effects, with hitchhiking extending across a 25cM region around the suppressor locus; linked variants were found to increase in frequency by up to 70% via hitchhiking. Immediately following the sweep there was a reduction in diversity at loci that co-segregated with suppression, and local LD. Some of the hitchhiking variants were novel in post-suppressor samples, indicating that the suppressor was introduced to the island via migration rather than evolving *de novo* within the population. At some loci secondary

alleles, that were often novel, were found to increase in frequency in 2010, even though the primary suppressor-linked allele remained at a stable frequency. This demonstrates that different alleles can come to hitchhike as association between the selected allele and linked variants is broken down over time. By 2010 some of the novel variants that seemingly hitchhiked into the population alongside the suppressor had shown a significant decline in frequency, possibly as a result of purifying selection.

This chapter extends the work undertaken by Hornett *et al*., (2014) to examine the changing signature of a selective sweep over time, by going on to also examine the signature of the selective sweep for the suppressor as it moved across space, travelling from Upolu and onto the neighbouring island of Savai'i. By looking at samples collected at three different time points on Savai'i: prior to the selective sweep (2001), immediately following the selective sweep (2007), and then again after the completion of the selective sweep (2010) and comparing them to samples collected at equivalent time points on Upolu (2001, 2006, 2010) it will be possible to assess how strength and breadth of the selective sweep changes over time and how this goes on to affect genetic differentiation in a structured population.

Methods

Samples

The samples used in this study had previously been collected from Upolu and Savai'i, Independent Samoa in 2001, 2006, 2007, and 2010 (table 4.1). Samples used in this study represent 3 time points across both islands: prior to the arrival of the suppressor (2001), just after the suppressor had spread across either Upolu (2006) or Savai'i (2007), and sometime after the spread of the suppressor (2010). Assuming that there are around 10 generations per year then this represents a period of time some 30-40 generations after the initial selective sweep for the suppressor. DNA was extracted from the Savai'i 2010 samples using the Qiagen DNeasy kit and following the manufacturer's protocol.

Population	Suppressor Status	N	Reference
Upolu, Independent Samoa (2001)	Absent	41	Dyson & Hurst 2004 Hornett *et al*., 2014
Upolu, Independent Samoa (2006)	Present	41	Charlat *et al*., 2007 Hornett *et al*., 2014
Upolu, Independent Samoa (2010)	Present	46	Hornett *et al*., 2014
Savai'i, Independent Samoa (2001)	Absent	23	Charlat *et al*., 2005
Savai'i, Independent Samoa (2007)	Present	15	Unpublished
Savai'i, Independent Samoa (2010)	Present	39	Unpublished

Table 4.1: *H. bolina* samples collected from Upolu and Savai'i, Independent Samoa at different time points between 2001 and 2010.

Marker amplification and sequencing

A total of 8 markers were selected that spanned an approximately 30cM region of *H. bolina* chromosome 25 surrounding the suppressor region (table 4.2). The markers Hb80, Hb89, Hb99, Hb131, Hb152, and Hb175 are derived from Hornett *et al*., 2014; the Hbdoublesex and HbProspero markers were developed as described in chapter 3. Samples derived from Upolu were amplified using only the Hbdoublesex and HbProspero markers, as they had previously been amplified and sequenced using the other 6 markers (Hornett et al. 2014), whilst all samples from Savai'i were amplified and then sequenced using the entire marker set.

PCR amplification was conducted in a total volume of 12 µl, comprising 4µl ddH_2O, 6µl of 2x GoTaq Hot Start Green Master Mix (Promega), 0.5pmol forward primer, 0.5pmol reverse primer and 1 µl of DNA template. The PCR cycling conditions consisted of an initial denaturation period of 1 minute at 95°C, followed by 35 cycles of 15 seconds at 95°C, 15 seconds at a marker-specific annealing temperature, then 72°C for a marker-specific extension time. This was followed by a final extension of 7 minutes at 72°C. The optimised annealing temperatures and extension times for each marker can be found in table 4.2.

Resulting PCR products were visualised on an agarose gel. Successfully amplified samples were purified and then sequenced using Sanger technology following the PCR product purification and sequencing protocol (appendix ii).

Marker	F 5'-3'	R 5'>3'	Ta (°C)	Extension time (sec)	Reference
Hb80	GTTCACGACAGCGTTGCTAT	CAATGCCAGCACACATTTCT	52	55	Hornett *et al*., (2014)
Hb89	GGCTGCAACGCTATAATGTG	TTTCCCTTTGGTAATGCTGTC	52	55	Hornett *et al*., (2014)
Hb99	TGCAAGATAACTACAATTTGGCTTT	TGCCACTAATGCAGCATCTC	53	30	Hornett *et al*., (2014)
Hb131	TCTGAAGAGGATGAAAAGAGTGG	GCGAATAAAGGCGTGTTCTG	53	30	Hornett *et al*., (2014)
Hb152	CCTCGTGCAGCAAATACTGA	AGTTCTTGGCGCATGACTTC	53	30	Hornett *et al*., (2014)
HB175	AACAGTGCAGCACAACTTGG	CCGGTTCCATTAGTGCTTTC	52	55	Hornett *et al*., (2014)
Hbdoublesex	GCGTCGCGGAAGATAGATG	GCTTCCTGGCGTACTCGTTGATG	53	30	This study
HbProspero	GGATAGCGGACCCAGAAGAACAT	CCACCGACTTGCCTTACGA	57	60	This study

Table 4.2: Primer and optimal PCR amplification conditions for markers used in this study.

Data Analysis

Sequences for each population were aligned using Geneious Pro v.5.6.6 (Kearse et al. 2012) and scored by eye for polymorphisms (SNPs, indels) that varied within and between populations. Allele haplotypes for each marker were reconstructed using PHASE v.2.1.1 (Stephens & Scheet 2005; Stephens et al. 2001), which was run with 1000 iterations, a thinning interval of 100 and 1000 burn-in iterations. Allele frequencies were calculated for each marker, across all time points for both Savai'i and Upolu. Alleles linked to the arrival of the suppressor were identified by an increase in their frequency in post-suppressor populations.

Nucleotide diversity and neutrality statistics (number of segregating sites (S), nucleotide diversity (π), θ per site, Tajima's D, number of haplotypes, and haplotype diversity (Hd)) were calculated for each marker, across all time points for both Upolu and Savai'i using MEGA v.7 (Kumar et al. 2016). Pairwise F_{ST} was calculated at each marker both between samples taken from the same island at different time points, and also for samples taken at the same time point but from different islands. An exact G test used to determine whether the populations were significantly differentiated in terms of allele frequency distributions. Pairwise $F_{ST,}$ exact G test, and also linkage disequilibrium (LD) were calculated using GENEPOP v4.2 (Raymond & Rousset 1995; Rousset 2008).

Results

Selective sweep for the suppressor on Savai'i

The selective sweep associated with the evolution of the suppression of *Wolbachia*-induced male-killing on Savai'i was found to affect an approximately 20cM region of *H. bolina* chromosome 25, extending from Hb80 at 0cM to Hb152 at 22.3cM. The sweep did not extend to Hb131 on Savai'i where the allele frequency distributions of post-suppressor populations were not found to significantly differ between pre-suppressor and post-suppressor populations.

The selective sweep peaked at the HbProspero marker (figures 4.1, 4.2). Here, the suppressor-linked haplotype reached a frequency level of 0.75 in 2007, which was maintained at 0.77 in 2010 (figure 4.2). This marker also experienced the greatest degree of genetic differentiation between pre-suppressor and post-suppressor populations (F_{ST} 2001-2007=0.332, F_{ST} 2001-2010=0.369; table 4.3, figure 4.1) and showed a reduction in π (table 4.4). At Hbdoublesex, the balanced equilibrium of two haplotypes that has been recorded in wild, suppressor-present populations was also seen post-sweep on Savai'i (chapter 3).

Selective sweep on Upolu vs sweep on Savai'i

Overall, the intensity of the selective sweep was greater on Upolu than on Savai'i. On Upolu, hitchhiking was found to extend from Hb80 (0cM) to Hb131 (30.6cM), whilst on Savai'i hitchhiking extended from Hb80 (0cM) to Hb152 (22.3cM).

The alleles that increased in frequency with the arrival of the suppressor onto Savai'i were found to be identical to those linked with the evolution of suppression on Upolu (figure 4.2). The extent to which these suppressor-linked alleles increased in frequency within post-sweep populations was found to be greater on Upolu than on Savai'i. At HbProspero, known to be located close to the suppressor (chapter 3), the suppressor-linked haplotype was novel to both islands. On Upolu it attained a frequency of 0.71 in 2006, rising to 0.96 in 2010, whilst on Savai'i it attained a frequency of 0.75 in 2007, rising to 0.77 in 2010.

Genetic differentiation between pre-sweep and post-sweep populations was higher on Upolu than Savai'i (table 4.3, figure 4.1). Post-suppressor populations were found to be significantly differentiated from the pre-suppressor population across all markers on Upolu. This was not recapitulated on Savai'i where post-suppressor populations were found to be significantly differentiated from the pre-suppressor population across all markers, save for Hb131 the marker most distally located from the suppressor region. Between 2006 and 2010 on Upolu significant changes were only found to occur at HbProspero, where an increase in the frequency of the suppressor-linked allele led to further population differentiation. Significant genetic differentiation between the 2007 and 2010 Savai'i samples was seen to occur at Hb152 and also at HbProspero although, unlike on Upolu, this wasn't associated with the continued increase in frequency of the primary suppressor-linked allele.

	Upolu			Savai'i		
Marker	**2001 x 2006**	**2001 x 2010**	**2006 x 2010**	**2001 x 2007**	**2001 x 2010**	**2007 x 2010**
HB80	0.054 ***	0.049 ***	-0.007	0.136***	0.031**	0.019
HB89	0.170 ***	0.163 ***	0.014	0.147***	0.110***	0.005
Hbdoublesex	0.183 ***	0.206 ***	0.000	0.236***	0.258***	0.001
HbProspero	0.318 ***	0.560 ***	0.128 ***	0.332***	0.369***	0.006*
HB99	0.204 ***	0.180 ***	-0.005	0.113**	0.117***	-0.024
HB175	0.190 ***	0.152 ***	0.001	0.076***	0.126***	0.000
HB152	0.149 ***	0.045 ***	0.026	0.084***	0.012**	0.012*
HB131	0.116 ***	0.135 ***	-0.009	0.062	0.020	-0.004

Table 4.3: Pairwise F_{ST} values between three time points on Upolu and Savai'i at 8 markers on *H. bolina* chromosome 25. Standardised genetic differentiation at 8 markers between populations from three time points: prior to the evolution of the suppressor, just after the evolution of the suppressor, and sometime after the evolution of the suppressor on Upolu (2001, 2006, 2010) and Savai'i (2001, 2007, 2010), Independent Samoa. Significantly differentiated populations are denoted by *** ($p<0.001$), ** ($p<0.01$), * ($p<0.05$).

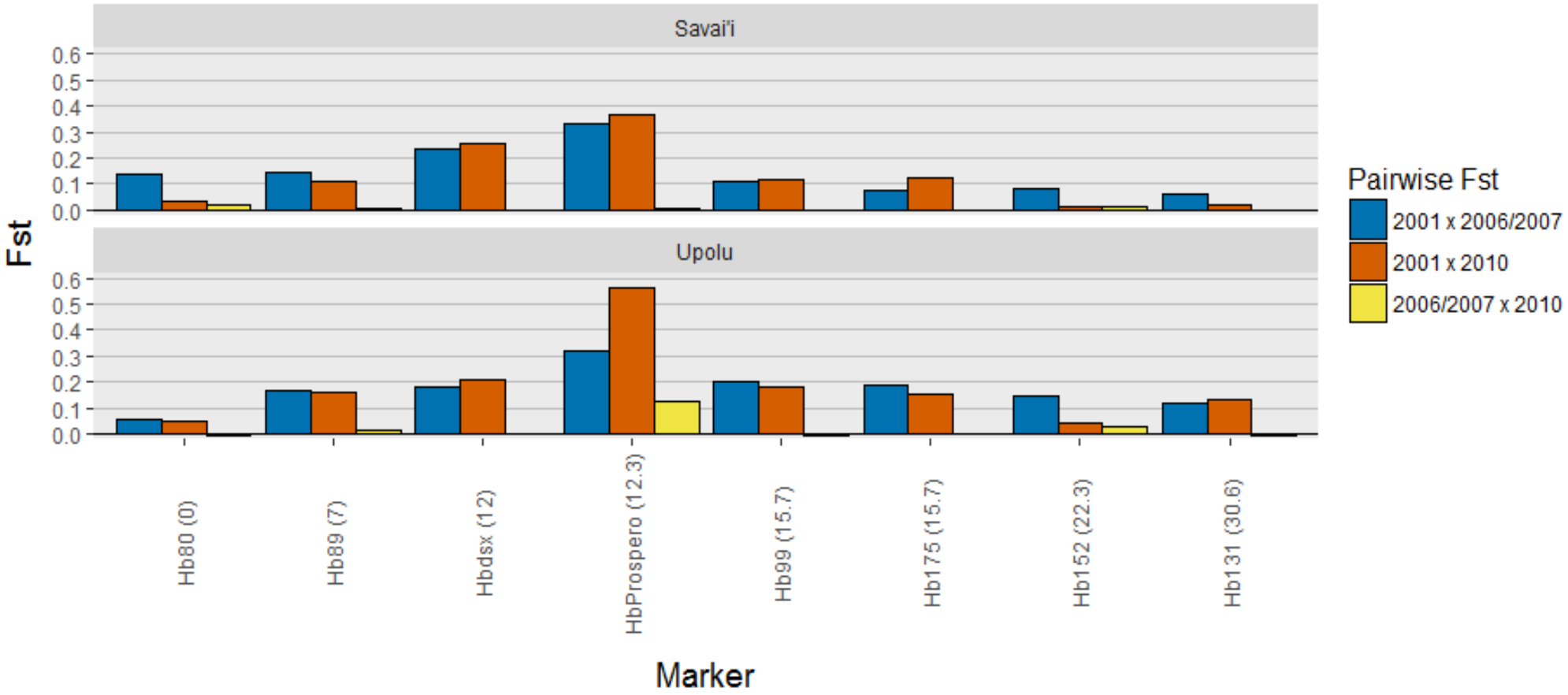

Figure 4.1: Pairwise F_{ST} at 8 markers on *H. bolina* chromosome 25, between samples from three time points on Upolu and Savai'i, Independent Samoa. Genetic differentiation between samples collected during 2001, 2006, 2010 on Upolu, and from 2001, 2007, and 2010 on Savai'i. Time points represent three different periods: prior to the arrival of the suppressor (2001), just as after the spread of the suppressor across each island (2006 Upolu, 2007 Savai'i), and sometime after the introduction of the suppressor (2010). Genomic distances of each marker are shown in shown in brackets and measured in cM.

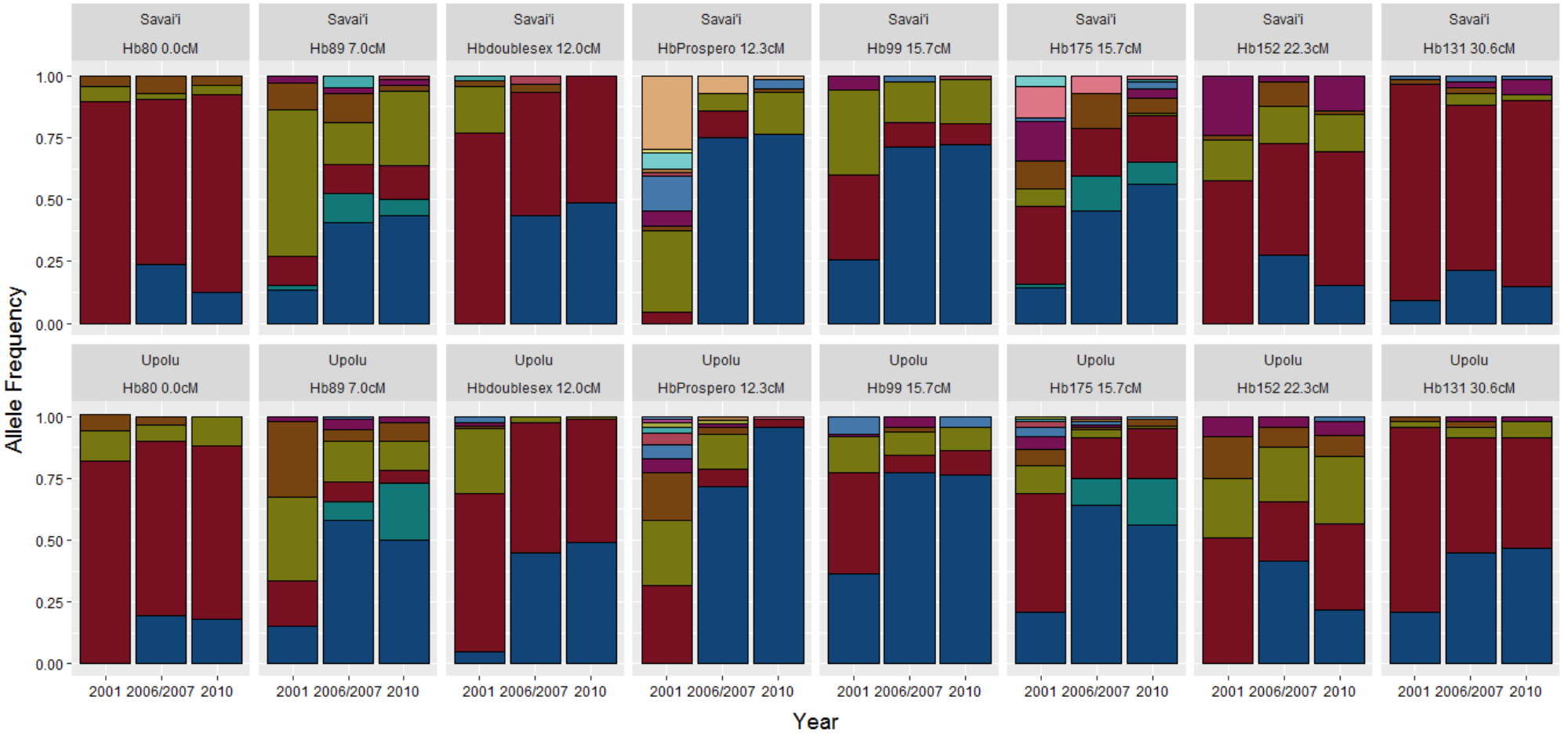

Figure 4.2: Change in allele frequencies across 8 markers on *H. bolina* chromosome 25 on Upolu and Savai'i, Independent Samoa. Allele frequencies of *H. bolina* collected from Upolu in 2001, 2006, 2010 and from Savai'i in 2001, 2007 and 2010. Allele frequency is shown on the y-axis and year of sample (2001, 2006/2007, 2010) along the x-axis. The major allele that increased in frequency with the introduction of the suppressor is coloured in red, secondary alleles that also increased in frequency are coloured in a lighter shade of red. Markers are ordered relative to their position along the chromosome, with position indicated in cM in the title of each plot.

There was an increase in LD associated with the evolution of suppression on Upolu in 2006, which had started to break down by 2010. On Savai'i however, there was no evidence of LD associated with the evolution of the suppressor in 2007 (figure 4.3).

Between 2001 and 2006 on Upolu, or 2007 on Savai'i, there was a reduction in nucleotide diversity (π) at markers linked close to the suppressor region (HbProspero, Hb99, Hb175). Contrastingly, there was an increase in π at markers located further from the suppressor region (Hb80, Hb89, Hb152, Hb131) on both Upolu and Savai'i (table 4.4).

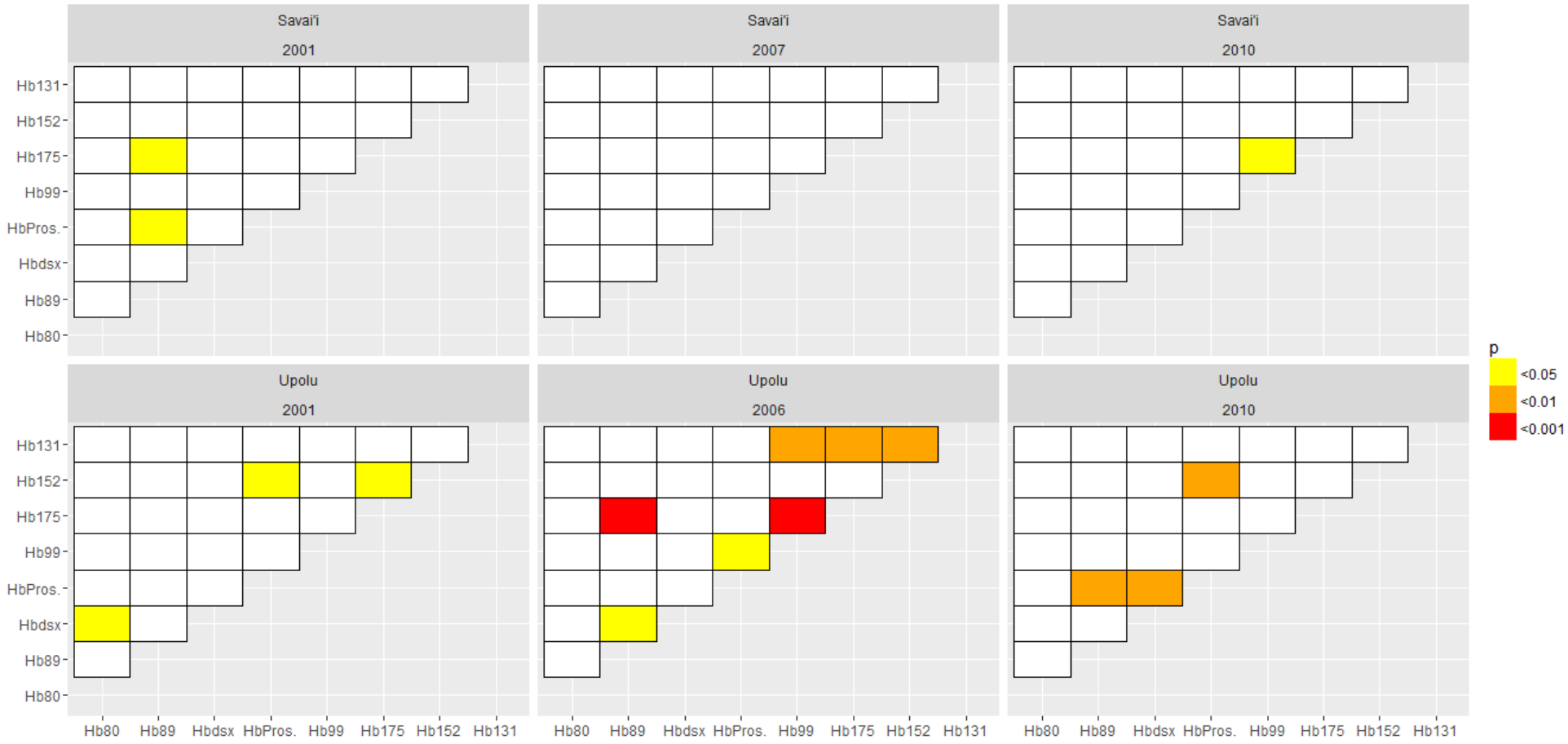

Figure 4.3: Linkage disequilibrium along *H. bolina* chromosome 25 at three time points for populations from Upolu (2001, 2006, 2010) and Savai'i (2001, 2007, 2010), Independent Samoa. Significance is denoted by colour: red ($p<0.001$), orange ($p<0.01$), yellow ($p<0.05$).

Marker	Island	Year	n	#sites	# segregating sites	π	θ (per site)	Tajima's D	# haplotypes	Hd	Reference
Hb80	Upolu	2001	92	173	2	0.00194	0.00227	-0.22493	3	0.320	Hornett *et al*., (2014)
		2006	88	173	3	0.00295	0.00343	-0.25893	4	0.466	Hornett *et al*., (2014)
		2010	84	173	2	0.00294	0.00231	0.43502	3	0.466	Hornett *et al*., (2014)
	Savai'i	2001	68	173	2	0.00114	0.00241	-0.87432	3	0.193	This study
		2007	42	173	3	0.00321	0.00403	-0.43872	4	0.505	This study
		2010	80	173	3	0.00213	0.00350	-0.73512	4	0.346	This study
Hb89	Upolu	2001	92	162	7	0.01325	0.00848	1.33454	5	0.744	Hornett *et al*., (2014)
		2006	90	162	8	0.01257	0.00974	0.71821	7	0.629	Hornett *et al*., (2014)
		2010	78	162	8	0.01200	0.01002	0.49701	6	0.683	Hornett *et al*., (2014)
	Savai'i	2001	66	212	8	0.00927	0.00800	0.40978	6	0.614	This study
		2007	42	212	7	0.01056	0.00778	0.97896	7	0.782	This study
		2010	80	212	7	0.01013	0.00676	1.21073	6	0.700	This study
Hbdoublesex	Upolu	2001	80	104	4	0.01683	0.00777	2.40841	6	0.528	This study
		2006	80	104	4	0.01504	0.00777	1.93449	3	0.528	This study
		2010	92	104	4	0.01479	0.00755	1.92916	3	0.516	This study
	Savai'i	2001	48	104	4	0.01294	0.00867	1.12887	4	0.378	This study
		2007	30	104	3	0.01481	0.00728	2.43322	4	0.579	This study
		2010	80	104	3	0.01460	0.00582	2.81827	2	0.506	This study
HbProspero	Upolu	2001	88	693	11	0.00422	0.00314	0.91005	10	0.792	This study
		2006	80	693	9	0.00225	0.00270	-0.43334	3	0.528	This study
		2010	92	693	6	0.00067	0.00170	-1.37929	3	0.516	This study
	Savai'i	2001	64	693	10	0.00285	0.00305	-0.18550	10	0.786	This study
		2007	28	693	6	0.00232	0.00223	0.12547	4	0.579	This study
		2010	60	693	6	0.00126	0.00186	-0.78680	5	0.389	This study
Hb99	Upolu	2001	96	244	4	0.00369	0.00319	0.31203	5	0.682	Hornett *et al*., (2014)
		2006	96	244	4	0.00217	0.00319	-0.63726	5	0.394	Hornett *et al*., (2014)
		2010	72	244	3	0.00179	0.00254	-0.56157	4	0.401	Hornett *et al*., (2014)
	Savai'i	2001	70	244	3	0.00432	0.00255	1.32955	4	0.706	This study
		2007	42	244	3	0.00208	0.00286	-0.58222	4	0.463	This study
		2010	68	244	3	0.00208	0.00257	-0.36967	4	0.448	This study

Marker	Island	Year	n	#sites	# segregating sites	π	θ (per site)	Tajima's D	# haplotypes	Hd	Reference
Hb175	Upolu	2001	96	517	9	0.00245	0.00339	-0.69233	9	0.712	Hornett *et al*., (2014)
		2006	92	517	9	0.00136	0.00342	-1.52328	9	0.555	Hornett *et al*., (2014)
		2010	80	517	6	0.00154	0.00234	-0.79808	6	0.615	Hornett *et al*., (2014)
	Savai'i	2001	70	517	8	0.00337	0.00321	0.12845	9	0.831	This study
		2007	42	517	5	0.00233	0.00225	0.09187	5	0.731	This study
		2010	80	517	9	0.00175	0.00352	-1.29619	9	0.642	This study
Hb152	Upolu	2001	96	312	3	0.00301	0.00187	1.10203	4	0.654	Hornett *et al*., (2014)
		2006	96	312	4	0.00379	0.00250	1.03468	5	0.720	Hornett *et al*., (2014)
		2010	92	312	5	0.00362	0.00252	0.88125	6	0.755	Hornett *et al*., (2014)
	Savai'i	2001	54	312	3	0.00291	0.00211	0.76929	3	0.595	This study
		2007	40	312	3	0.00321	0.00226	0.91083	4	0.692	This study
		2010	78	312	4	0.00310	0.00260	0.39492	5	0.651	This study
Hb131	Upolu	2001	96	279	3	0.00149	0.00209	-0.52085	4	0.397	Hornett *et al*., (2014)
		2006	94	279	4	0.00239	0.00280	-0.29678	5	0.585	Hornett *et al*., (2014)
		2010	92	279	3	0.00240	0.00211	0.24944	4	0.585	Hornett *et al*., (2014)
	Savai'i	2001	64	279	4	0.00095	0.00303	-1.47623	4	0.229	This study
		2007	42	279	5	0.00224	0.00416	-1.16085	6	0.518	This study
		2010	80	279	4	0.00170	0.00289	-0.85483	5	0.416	This study

Table 4.4: Measures of polymorphism and neutrality at 8 *H. bolina* chromosome 25 markers for populations from Upolu (2001, 2006, 2010) and Savai'i (2001, 2007, 2010). Information listed is as follows: marker, island of origin, sample year, number of alleles with reconstructed haplotypes in each sample (n), number of sites in bp (#sites), number of segregating sites (# segregating sites), nucleotide diversity (π), θ (per site), Tajima's D, number of haplotypes present in each sample, haplotype diversity (Hd), and reference for each data set.

Chapter three proposed the hypothesis that the evolution of suppression is associated with duplication of *doublesex*. Changes associated with suppression at Hbdoublesex were very similar on both Upolu and Savai'i. As found in other, wild *H. bolina* suppressor-present populations; post-sweep the Upolu and Savai'i populations both moved towards having a balanced equilibrium of two alleles at this locus. This was associated with an increase π in on Savai'i but a decrease in π on Upolu (table 4.4). After HbProspero, Hbdoublesex showed the greatest genetic differentiation between pre-sweep and post-sweep populations (table 4.3).

Prior to the selective sweep for the suppressor, there was little genetic differentiation within *H. bolina* from across Independent Samoa. Immediately following the spread of the suppressor, genetic differentiation was found to disappear entirely at most of the markers (Hb80, Hb89, Hb99, Hb131, Hb152, Hbdoublesex, HbProspero; figure 4.4). By 2010, aside from HbProspero, there were still no significant levels of genetic differentiation between Upolu and Savai'i samples at markers that co-segregate with the suppressor (Hb99, Hb175, Hbdoublesex; figure 4.4). Significant levels of genetic differentiation were found once again at other markers (Hb80, Hb89, Hb131, Hb152, HbProspero; figure 4.4). At Hb131, the marker most distal from the suppressor region, genetic differentiation between Upolu and Savai' was five times higher in it was prior to the sweep for the suppressor (2001: F_{ST}=0.031, 2010: F_{ST}=0.154; figure 4.4).

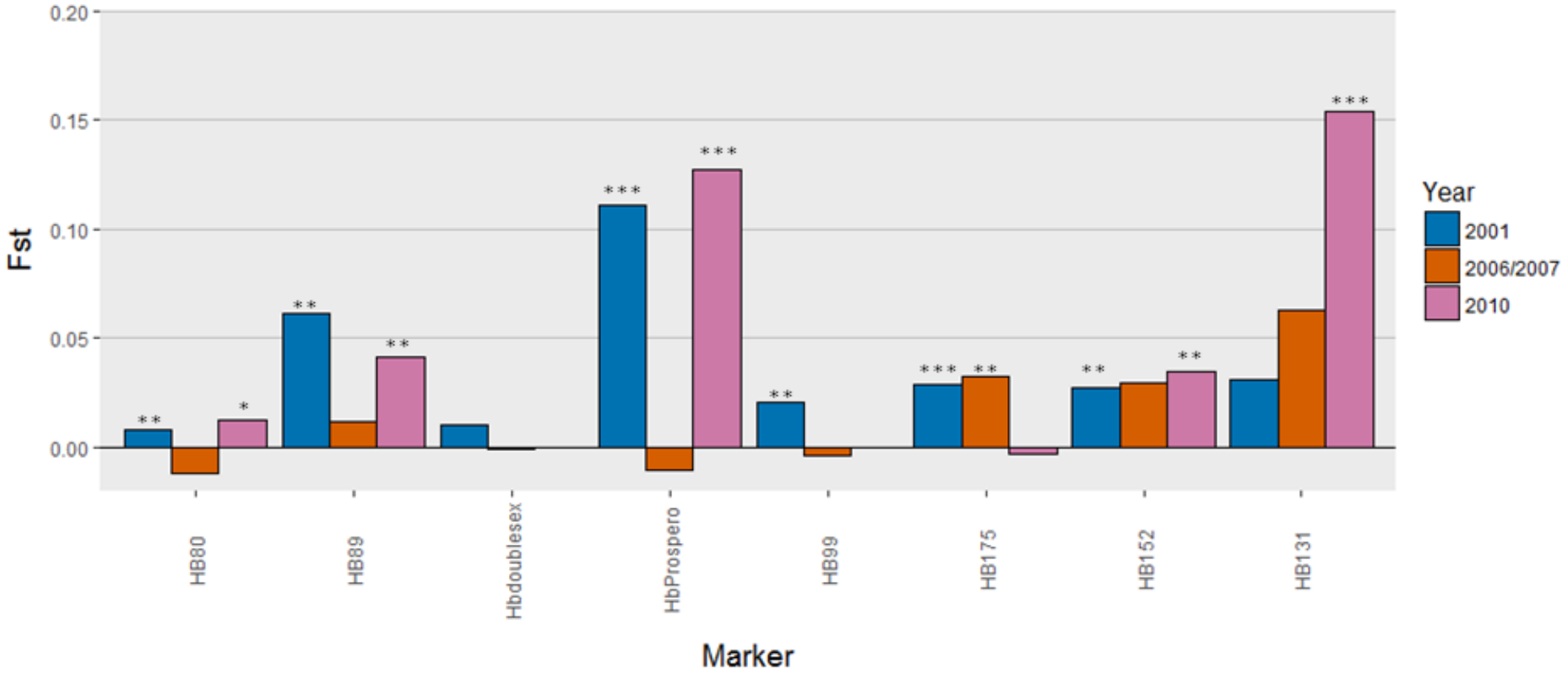

Figure 4.4: Genetic differentiation between populations of *H. bolina* on Upolu and Savai'i at three time points for 8 chromosome 25 markers. Pairwise F_{ST} between *H. bolina* populations from Upolu and Savai'i pre-sweep (2001), immediately post-sweep (2006 Upolu vs. 2007 Savai'i) and sometime following the sweep (2010). Significant deviations from F_{ST}=0 are shown by ***(p<0.001), **(p<0.01), *(p<0.05).

Discussion

Extremely female-biased sex ratios are known to have existed for at least a century amongst the *H. bolina* population of Independent Samoa with, sex ratio bias associated with high prevalence of the male-killing strain of *Wolbachia w*Bol1 (Dyson & Hurst 2004; Dyson et al. 2002). However, in May 2005 males were spotted more frequently. Formal investigations revealed that, like their South-East Asian counterparts, butterflies on Samoa had evolved the ability to suppress the male-killing effects of *Wolbachia*. Selection on the suppressor was incredibly strong, so much so that by 2005/6 the sex ratio on Upolu was near parity, followed by the return of the sex ratio to parity on Savai'i in 2006/7 (Charlat et al. 2007b). This selective sweep left a very broad and strong signature of selection on the *H. bolina* population of Upolu (Hornett et al. 2014). This chapter analysed the selective sweep as it travelled in space from Upolu and onto Savai'i.

The selective sweep on Savai'i was narrower than on Upolu. Whilst hitchhiking extended across a 30cM region on Upolu (Hornett *et al*., (2014) cite 25cM but the updated linkage map in chapter 3 maps the markers that delimit the breadth of the sweep (Hb80, Hb131) as being 30cM apart), it was only shown to extend 22cM on Savai'i. This is in accord with theory that predicts that hitchhiking around the locus under selection should narrow, as recombination breaks down association as the selective sweep travels through space (Barton 2000; Slatkin & Wiehe 1998).

The Savai'i selective sweep was also less intense than the Upolu sweep. At HbProspero, the locus that is closest to Hbdoublesex, the hypothesised target of selection (chapter 3), a novel allele appears on Upolu in 2006 at a frequency of 0.71 rising to 0.96 in 2010. On Savai'i this allele is also novel in the 2007 population and has seemingly hitchhiked to a frequency of 0.77, which is maintained in the 2010 sample. The primary hitchhiking allele at other markers attained similar frequencies on Savai'i as on Upolu. At Hb152 and Hb131, the markers furthest away from the suppressor locus, the hitchhiked variant attained lower frequencies on Savai'i than on Upolu. This again shows that association between the locus under selection and neutral alleles is broken down as a sweep progresses across space and time, resulting in a weaker sweep.

Although the selective sweep for the suppressor was weaker on Savai'i than it was on Upolu, it still bears the signature of a very intense selection event. At all markers examined, the allele that hitchhiked in association with suppression on Upolu was also found to be the hitchhiking variant on Savai'i. This is the signature of a sweep that occurred rapidly, so that association between the suppressor allele and initial linked variants had not fully broken down before the sweep moved onto Savai'i. The hitchhiked variants that were novel in the suppressor-present populations of Upolu were also novel on Savai'i. This is consistent with the islands being within close proximity of each other, and with *H. bolina* being a dispersing species, as migratory butterflies containing the suppressor were able to travel from Upolu to Savai'i before association between hitchhiking alleles and the suppressor was broken down.

Immediately following the selective sweep, genetic differentiation between *H. bolina* from Upolu and Savai'i disappeared completely at markers spanning a 30cM range. This may in part be attributed to the smaller sample size for 2007, relative to other samples. By 2010 genetic differentiation had increased at markers that did not co-segregate with suppression. This is consistent with the idea that alleles coalesce with those that were originally at the wavefront of selection, which reduces differentiation (Barton et al. 2013). At Hb131 genetic differentiation in the 2010 suppressor-present sample, had increased five-fold relative to the pre-suppressor sample. This is the marker furthest away from the suppressor locus, so recombination is likely to break down associations between neutral alleles and the suppressor much more rapidly than at other markers, such that there was no associated allele at the point the suppressor arrived in Savai'i. This is consistent with theory which predicts that as linkage breaks down different alleles will hitchhike with the selected allele as the sweep travels through space, which may in turn result in an increase in genetic differentiation across a subdivided population (Bierne 2010; Kim 2013; Slatkin & Wiehe 1998).

Sex-biased populations, such as that of Independent Samoa have small effective population sizes (Ne) with Ne being smaller than the number of breeding individuals within the population (N). The greater the sex bias then the smaller the Ne (Wright, 1931). The extreme sex bias seen on Independent Samoa predicts a population with an Ne much smaller than N; such populations are more susceptible to the vagaries of genetic drift –

fixation of both positively and negatively selected variants being more likely to occur. *H. bolina* is known to be migratory and it is thought that the *H.bolina* population of Independent Samoa has been sustained, in spite of the extremely female-biased sex ratio due to migratory individuals which would also have increased the Ne of Independent Samoan populations. It would be expected that prior to the introduction of the suppressor Ne and genetic diversity would have been broadly uniform across the *H. bolina* genome. It would be expected that after the selective sweep, Ne and genetic differentiation would decrease and the population as a whole would become less genetically differentiated at markers linked to the locus under selection. Immediately following the selective sweep for the suppressor a reduction in diversity on both Upolu and Savai'i was seen and the two populations become more genetically similar, as shown in this chapter and Hornett *et al*., (2014). Such changes are however not predicted to occur uniformly across the genome. At markers unlinked to the suppressor, Hornett et al., (2014) did not find any significant changes in genetic diversity or differentiation immediately following the selective sweep. As the suppressor was introduced via migration and selection would have been very strong in the initial stages of the sweep then genetic material from this migratory individual would have entered the population thus increasing Ne, diversity and making the population more similar genetically, to other suppressor populations. However, as there is no selective pressure or hitchhiking effect at such unlinked loci the introduced genetic material would have been broken down by recombination making it detectable only in sample sizes much larger than those used in this study of the Samoan selective sweep. Future work should assess the changes in genetic diversity and differentiation associated with the introduction of the suppressor at unlinked markers using the Savai'i population.

Chapter three proposed the hypothesis that *doublesex* is the target of selection and that the evolution of suppression is associated with duplication of *doublesex*. The selective sweep on Savai'i is consistent with this hypothesis. The focal point of the selective sweep on Savai'i was found to correspond with the genomic region containing the suppressor, as it was on Upolu. That the suppressor-linked allele at *Prospero* did not approach fixation on Savai'i, as it did on Upolu, clarifies that although *Prospero* is closely linked to the suppressor locus, it is not itself the locus under selection. Furthermore, Savai'i samples collected after the introduction of the suppressor were also found to have an excess of apparent heterozygotes

at the Hbdoublesex marker, a phenomenon that is found in other *H. bolina* samples from populations in which the suppressor is present (chapter three). A similar pattern of an excess of apparent heterozygotes after a selection event has also seen to result from gene duplication in *Culex pipens* (Lenormand et al. 1998). If duplication of *doublesex* has occurred then the presence of only two haplotypes in the 2010 Savai'i sample indicates that the suppressor has reached fixation between 2007 and 2010 and that the sweep is complete.

Future work should involve further sampling and examination of the selective sweep in greater detail using NGS technologies. If a good reference assembly of *H. bolina* chromosome 25 was created then the Savai'i and Upolu samples from each time point could be pooled, sequenced and aligned to the reference assembly. This would permit a sliding window analysis, which would show more clearly the locus of selection, how far hitchhiking extended along the chromosome, how hitchhiking broke down as the sweep progressed in both time and space, how purifying selection affects the shape of the sweep and the extent to which the sweep affected genetic differentiation along the chromosome. If samples were collected from across Independent Samoa at future time points, and also from South-East Asian populations, where the suppressor has been present for much longer, then this would reveal more about the signature of selective sweeps as they travel across greater scales of space and time.

This chapter characterised a selective sweep for a mutation that suppresses the male-killing effects of a strain of *Wolbachia* in *H. bolina* on Independent Samoa, as it progressed across both space and time. The impact of the sweep was both intense and broad but the sweep became weaker as it progressed from the island of Upolu and onto the neighbouring island of Savai'i. This is consistent with theory which predicts that sweeps should narrow over space and time, as recombination breaks down linked alleles. The signature of selection on Savai'i is also consistent with the hypothesis that the evolution of suppression is associated with duplication of *doublesex*. It is extremely rare to catch evolution in action and to have samples from before and after a selection event to use for comparison. As such, the insight gained from studying this system could have wider reaching consequences by furthering knowledge of the signature of positive selection, which can be used to improve models of selection events and methods for detecting positive selection.

Chapter 5

Is there a hotspot for the evolution of melanism in Lepidoptera?

Abstract

Melanic forms have evolved in numerous species of Lepidoptera. The loci that control melanic forms of the peppered moth *Biston betularia* and the silkworm *Bombyx mori* have been mapped to a region of the lepidopteran genome that also controls aspects of wing patterning in other species. This has led to the hypothesis that this region is a 'hotspot' for the evolution of adaptive traits of wing patterning in Lepidoptera. This chapter investigates whether this hotspot is also implicated in the control of a melanic form of the pale brindled beauty *Phigalia pilosaria*, and of 3 melanic strains of the Mediterranean flour moth *E. kühniella*, the former representing a natural polymorphism and the latter laboratory-derived mutations. The wing patterning hotspot was found to control melanism in *P. pilosaria* but not in *E. kühniella*. The discovery that the melanic morph of *P. pilosaria* is also controlled by this region gives further credence to this genomic region being a hotspot for adaptation. That this region was not found to control melanism in *E. kühniella* indicates that there are other mutational possibilities that give rise to this phenotype outside of the 'hotspot' region.

Introduction

The convergent evolution of melanism in Lepidoptera is a good model system in which to investigate whether convergence at the phenotypic level goes hand in hand with convergence at the molecular level. Melanism refers to an organism "having dark or black pigmentation" (Majerus 1998, p. 291). This colour form variation can take the form of discrete melanic and non-melanic (typical) morphs, or a continuous range of morphs with varying degrees of melanic pigmentation (Majerus 1998; True 2003) . Melanic forms are found in many species of Lepidoptera (Majerus 1998), the adaptive significance of which has been extensively studied. Furthermore, there is an expanding range of genomic and developmental resources available for the study of lepidopteran systems.

Melanism famously arose in the peppered moth *Biston betularia* as an adaptive response to industrial pollution, termed industrial melanism. There are three *B. betularia* morphs: a dappled *typica* morph, a black *carbonaria* morph and an intermediate *insularia* morph (Clarke & Sheppard 1964). The *carbonaria* morph emerged in Greater Manchester during the 18th Century. It was more cryptic and successful than the *typica* morph in urban settings and rapidly usurped the typical morph across the industrial heartlands of the UK (Kettlewell 1973). It reached near-fixation levels before receding in frequency after the introduction of the Clean Air Acts of 1956 and 1968, in response to the changing selection pressure from a post-industrial environment (Cook 2003; Cook et al. 2002; Saccheri et al. 2008). Melanic forms are known to have a number of evolutionarily-adaptive values such as crypsis (Kettlewell 1973; Zink & Remsen 1986), mimicry (Clarke & Sheppard 1959), bacterial resistance (Goldstein et al. 2004; Gunderson et al. 2008), thermoregulation (Talloen et al. 2004; Walsberg 1983), defence against fungal attack (Chu et al. 2015), and wound healing (Sugumaran 2009).

The development of pigmentation in Lepidoptera occurs in two stages. First, patterning genes regulate the location of pigments in the insect body during development and then effector genes regulate the synthesis of these pigments. Pigments develop shortly before adult emergence. They are synthesised in epidermal cells, which produce scales of a single colour, and are then assimilated into the exoskeleton (Wittkopp & Beldade 2009; Wittkopp et al. 2003a; Wittkopp et al. 2003b). Colour pigments develop first and then melanic

pigments are laid down in the remaining unpigmented scales (Koch et al. 2000b; Koch et al. 1998).

Melanic pigments are produced in the melanin synthesis pathway (figure 5.1), which has been described in a number of insects (Futahashi & Fujiwara 2005; Futahashi & Fujiwara 2007; Gorman et al. 2007a; Gorman et al. 2007b; Wright 1987). As with the general insect melanisation pathway, two melanin pigments are produced in this pathway in lepidoptera: brown melanin, which has dopamine (dihydroxyphenylethylamine) as a precursor and black melanin, which has dopa (dihydroxyphenylalanine) as its precursor. Yellow sclerotin (NBAD sclerotin) and clear sclerotin (NADA sclerotin) pigments are also produced from the dopa and dopamine pigmentation pathways. The *Drosophila* gene *yellow* controls the dopa protein that is upregulated in cells that produce black dopa pigment, conversely *ebony* suppress melanin formation in cells where it is upregulated (Ferguson et al. 2011; Futahashi et al. 2008; Wittkopp et al. 2002a; Wittkopp et al. 2002b). In areas with dopamine melanin pigment, there is evidence that *tan* is upregulated (Ferguson et al. 2011; True 2003).

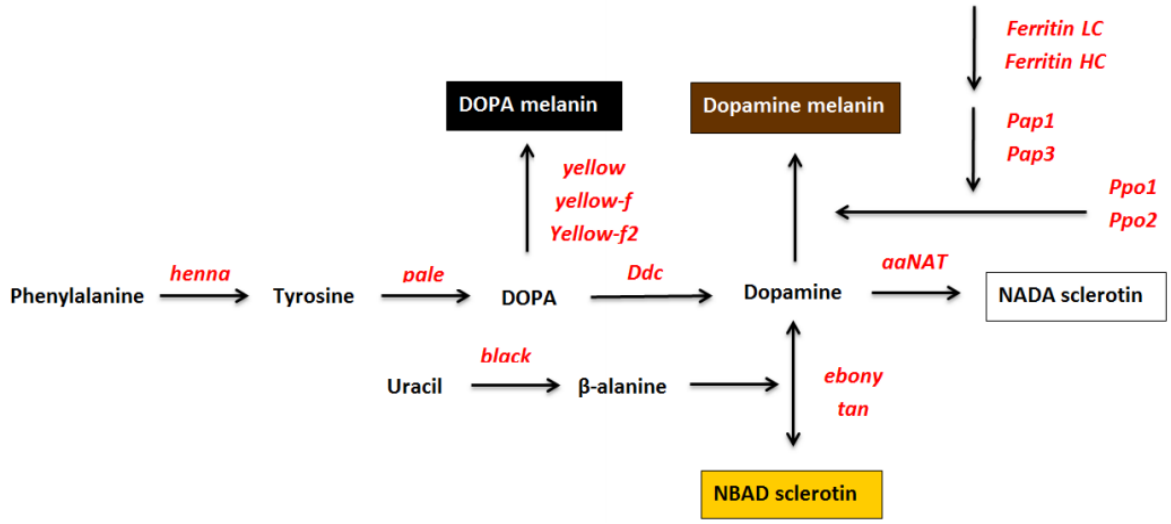

Figure 5.1: melanin synthesis pathway. Pigment precursors are in black. Genes involved in the pathway are in red italics. Pigments produced in the pathway are shown in correspondingly coloured boxes.

The molecular mechanisms that control melanic morphs have been investigated in a number of Lepidoptera species and were found to involve changes in the melanin synthesis pathway. The sex-linked melanic female form of *Papilio glaucus* is controlled by a single Y-linked gene; melanisation is brought about by suppression of *ebony* and reduced expression of the melanin precursor-producing gene *dopa decarboxylase* (Ddc) (Koch et al. 2000a; Koch et al. 2000b; Koch et al. 1998; Scriber et al. 1996). The melanic larval morph of *Manduca sexta* is, like *P. glaucus*, associated with increased DDC expression, and is also associated with a reduction in juvenile hormone (Hiruma & Riddiford 2009; Truman et al. 1973). Increased expression of both DDC and tyrosine hydroxylase were found to be responsible for melansim in a strain of *Spodoptera exigua* (Liu et al. 2015).

There are a number of melanic strains of the genomic model lepdiopteran *B. mori*. Deleterious changes within the *ebony* gene were found to control larval and pupal melanism found in the *sooty* strain (Futahashi et al. 2008). The *B. mori melanism* (*mel*) mutant was found to be controlled by mutational changes within the arylalkylamine-N-acetyltransferase gene *Bm-iAANAT*, which lower its expression levels, leading to an increase in dopamine which in turn results in the melanic pigmentation seen in *mel* mutants (Dai et al. 2010; Zhan et al. 2010). The locus controlling the Black moth *Bm* strain has been mapped to the same genomic region that controls *carbonaria* in *B. betularia* (Ito et al. 2016).

Molecular changes responsible for controlling melanic morphs and strains in the examples discussed above have, perhaps unsurprisingly, largely been found to have occurred within the melanin synthesis cascade. However, this is not the case for *B. betularia*, where no association was found between the *carbonaria* morph and genes recognised as being involved in melanin synthesis (van't Hof & Saccheri 2010). Rather, the mechanism that controls *carbonaria* has been localised to a 200kbp region of chromosome 17 (van't Hof et al. 2011). The orthologous genomic region has been implicated in the control of wing patterning in other Lepidoptera. The loci controlling the melanic *B. mori* Black moth strain and also the *B. mori* Wild wing spot *Ws* strain have both been mapped to this region (Ito et al. 2016). Changes within this region have also been found to control the Bigeye phenotype in *B. anynana* (Beldade et al. 2009; Saenko et al. 2010). Aspects of colour patterning in *Heliconius* butterflies have also been found to be controlled by changes at this locus (Kronforst et al. 2006; Papa et al. 2008). This region is also syntenic to the *H. melpomene*

supergene HmYb-HmSb-HmN, which controls wing pattern elements (Joron et al. 2006), the *H. numata P* locus which controls mimetic forms (Joron et al. 2011; Joron et al. 2006) and also to other colour patterns in *Heliconius* species (Baxter et al. 2008; Chamberlain et al. 2009; Ferguson et al. 2010; Merrill et al. 2011). The existence of these common, underlying mechanisms implies that a deeply conserved genetic mechanism for wing pattern adaptation could exist in this part of the lepidopteran genome and function as a hotspot for evolutionary adaptive mutations.

This chapter investigates whether the convergent evolution of melanic morphs in Lepidoptera is controlled by a shared genetic mechanism both within and between species. It does this by examining whether the genomic region that controls melanism in *B. betularia* also controls melanism in the pale brindled beauty *Phigalia pilosaria* and in three melanic strains of the Mediterranean flour moth *Ephestia kühniella.*

The pale brindled beauty *Phigalia pilosaria* is a geometrid moth with melanic morphs. There are two melanic forms of *P. pilosaria* (Figure 5.2), which are controlled by alleles at single locus (Lees 1974). The *monacharia* form is completely dominant to the somewhat lighter and patterned 'intermediate' form, which is itself completely dominant over the recessive non-melanic typical form (Lees 1974).

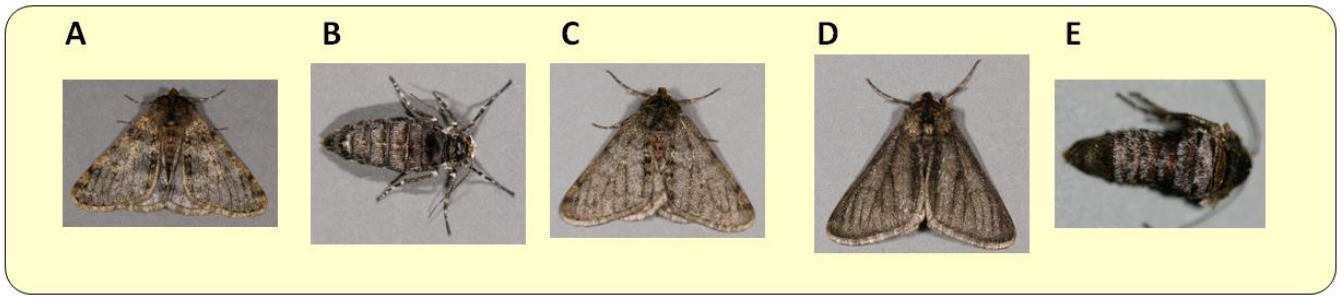

Figure 5.2: *Phigalia pilosaria* forms. Dorsal view of male and female *P. pilosaria* morphs: A. Male *typica* form, B. female *typica* form, C. male intermediate form, D. male *monacharia* form, E. female *monacharia* form.

The Mediterranean flour moth *Ephestia kühniella* is a member of the pyralidae family. Although melanic morphs are not found in the wild, melanic strains have emerged and been maintained in a lab environment. These strains include:

Ala nigra: An autosomal dominant mutant (Cotter 1974)
Mel: An autosomal recessive mutant (Marec 1991)
Abt: A strain with an autosomal recessive 'black' mutant (Kuhn & Henke 1929)

This chapter seeks to establish whether melanism in *P. pilosaria* and in three melanic strains of *E. kühniella* is controlled by the Lepidoptera wing patterning hotspot region. This is investigated first by creating mapping families for each species. Markers are then developed for use with these mapping families based on genes located on *B. betularia* chromosome 17, the chromosome that contains locus that controls the *carbonaria* morph. Further markers are developed for use with the *E. kühniella* mapping families to investigate whether the locus that controls selection in these strains is located within a linkage group that has been found to contain genes that have been found to be involved in the insect melanisation pathway in other Lepidoptera. Segregation analysis is used to determine whether or not a marker is associated with melanism.

Methods

Experimental design

Segregation analysis was used to map the region that controls melanism in *P. pilosaria* and *E. kühniella* using mapping families that were created using a reciprocal backcross design.

As discussed in chapters 2 & 3, crossing over during gamete formation occurs in male but not in female Lepidoptera (Maeda 1939; Suomalai.E et al. 1973; Traut 1977; Turner & Sheppard 1975). The result of this is that the maternally-derived chromosome is inherited completely and all of the alleles on that chromosome remain in linkage. The paternally-derived chromosome however did undergo recombination and so not all of the alleles on a chromosome remain linked.

A female-informative mapping family is created by crossing a female, heterozygous for melanism, to a homozygous male. The linkage group that the locus controlling the melanic phenotype belongs to can be determined using the female-informative family: if a marker is in the same linkage group as the locus that controls the melanic phenotype, then the maternal allele will co-segregate with melanic offspring.

The reciprocal male-informative mapping family is created by crossing a male, heterozygous for melanism, to a homozygous female. Co-segregation of the paternal allele with melanic offspring can be used to map the locus that controls melanism within a linkage group.

Markers located in regions syntenic to *B. betularia* chromosome 17, where the *carbonaria* locus is located, were first assessed for co-segregation with melanism. If this returned a negative result, markers on chromosomes syntenic to those on which canonical melanisation genes have previously been identified in other Lepidoptera species were then targeted.

Creating a mapping family for P. pilosaria

Segregation analysis was used to map the genomic region that controls the *monacharia* phenotype in *P. pilosaria*. A male-informative mapping family was created over the course of three seasons.

In the first season *P. pilosaria* eggs that resulted from a mating between a typical morph male were received from a commercial Lepidoptera supplier and raised in the lab. In the second season F1 typical morph virgin females were mated with wild-caught *monacharia* males (samples obtained from Cuerden Valley Park, Lancashire, courtesy of Graham Jones). The next season, the resulting F2 offspring were scored for sex and phenotype upon eclosion and then frozen at -80˚C for genetic analysis.

The *P. pilosaria* mapping family consisted of 32 offspring: 17 *monacharia* individuals (10 male, 7 female) and 15 *typica* individuals (6 male, 9 female).

Although the phenotype of the female parent of this cross is unknown, as the male parent was of the *typica* phenotype and the resulting F1 generation were all *typica* morphs (n=53), it is safe to assume that the female parent was also a *typica* morph. Likewise, it was not known whether the wild-caught F1 male was homozygous or heterozygous at the *monacharia*. As the proportion of *monacharia*:*typica* morphs of the F2 offspring (n= 17, n=16 respectively) did not differ significantly from the expected 1:1ratio for a homozygous typical x heterozygous *monacharia* cross (Pearson χ^2=0.030, d.f.=1, n=31, p=0.862) it can be presumed that the male was heterozygous at the *monacharia* locus.

Crosses were performed by placing a single male and female adult together in a mesh cage at dusk. Pairs were observed frequently at regular intervals to accurately ascertain whether or not coitus occurred. Mating pairs were left together until morning when the male was removed and the female transferred to a container in which she could lay her eggs. Once the female had finished laying, adults were stored at -80°C. Although *P. pilosaria* oviposit at the start of the year, their eggs do not hatch until spring, so eggs were kept under controlled temperature and light conditions until food became available. Upon emergence, larvae were raised on a diet of hawthorn leaves. Once caterpillars had successfully reached pupation they were stored in a cold room until it was time for them to emerge the following spring.

Creating mapping families for Ephestia kühniella

A reciprocal backcross design was used to create both male-informative and female-informative mapping families for the three *E. kühniella* strains. To create mapping families for the *Ala nigra* strain a homozygous wild-type (WT-C) female was crossed to a homozygous *Ala nigra* male, the female informative family was created by crossing an F_1 heterozygous female with a homozygous WT-C male. The *abt* strain mapping families were created by crossing a homozygous WT-C female with a homozygous *abt* male. The female informative family was created by backcrossing a heterozygous F_1 female with a homozygous *abt* male. The *wa* mapping family was created by crossing a homozygous *wa* female to a homoxygous *mel* male. The female mapping family was created by backcrossing a heterozygous F_1 female to a homozygous *mel* male (appendix v).

Parental crosses were performed by the Marec Lab (Academy of Sciences of the Czech Republic) and F_1 offspring sent to Liverpool as eggs. Larvae were raised on a diet of bran, honey and yeast. To ensure that crosses were only performed with virgin adults, males and females were separated as 5th instar larvae. At this stage of development it is possible to discern males due to the visibility of their reproductive organs. This however was not possible for the *wa* strain, where the sexes were instead separated upon eclosion; only females that displayed behavioural signs of being virgins, namely that they raised their abdomens in the air, were used for crosses.

F_1 crosses were performed by placing a single virgin female into a tub with a number of virgin males. Once copulation had been instigated, the mating pair was moved into a separate container which the female could lay her eggs. Upon emergence, F_2 progeny were scored by sex and phenotype before being stored at -20°C.

Marker development

A mixed methods approach was used to develop markers for the study. Markers were initially developed based on sequence generated by amplifying cDNA with degenerate primers (Fizzy in *E. kühniella*) and species-specific sequence already available through NCBI (GAPDH and wingless markers in *E. kühniella*). However markers were later designed using

new genomic resources as and when they became available. Namely, these were a whole genome fragment library for *P. pilosaria* and RNA-seq data for the *E. kühniella* sister species *Ephestia cautella* (Accession number PRJNA254197). The degenerate primers used in this study were designed by A. van't Hof. Primers for the E. *kühniella* markers Attracatin, Simkin, and 136 were designed by Mandy Lingbeek. All other markers were designed by myself.

As there is a high level of synteny in Lepidoptera (Pringle et al. 2007) marker location and intron/exon boundaries were estimated based on *B. mori*. A tBLASTx of potential markers was performed against the Silk Worm genome (Xia et al. 2004) using the Silkworm Genome Database (SilkDB) (Duan et al. 2010; Wang et al. 2005) to ensure that markers were not developed in genes with paralogues. Where possible, markers were designed across intronic regions, as they are more likely than exons to contain polymorphisms that can be used in mapping. Primers were designed with the aid of Oligo 6 and ordered from Eurofins.

The template used for amplification with degenerate primers was cDNA. cDNA was created following the 1st strand cDNA synthesis protocol (appendix iv) using RNA extracted with TRIZOL and following the RNA extraction protocol (appendix iii). Genomic DNA was used for amplification with species-specific primers. DNA was extracted by using either a whole *E. kühniella*, or a piece of tissue approximately 5mm3 from the abdomen of each *P. Pilosaria* sample. DNA was extracted following the DNA extraction protocol (appendix i).

Marker amplification and sequencing

Degenerate primers: cDNA was used as a template during PCR amplification with degenerate primers. Each PCR reaction took place in a total volume of 15µl consisting of 1.5 µl of 10x PCR Buffer I (Applied Biosystems), 0.3 µl dNTPs, 0.5 µl 10pmol forward primer, 0.5 µl 10pmol reverse primer, 0.12 µl Amplitaq GOLD (Applied Biosystems), 11.58 µl of ddH2O and 0.5 µl of cDNA. PCR cycling conditions were 9 minutes at 94°C followed by 40 cycles of 94°C for 30sec, 57°C for 30 sec, 72°C for 60 secs and finally 10°C for 5mins. PCR products were then resolved on an agarose gel electrophoresis to assess quality.

Species-specific primers: Amplification of species-specific primers was conducted in a total volume of 12 µl consisting of 4 µl ddH2O, 6µl of 2x GoTaq Hot Start Green Master Mix

(Promega), 0.5pmol of forward primer, 0.5pmol reverse primer and 1 µl of DNA. PCR cycling conditions consisted of an initial denaturation period of 1 minute at 95°C, followed by 35 cycles of 15 seconds at 95°C, 15 seconds at annealing temperature, extension time at 72°C, this was followed by a final extension of 7 minutes at 72°C. Optimal annealing temperatures (Ta) and extension times for each primer pair can be found in table 5.1.

Resulting PCR products were visualised on an agarose gel. Samples that had successfully been amplified were then purified and sequenced using Sanger technology following the PCR product and sequencing protocol (appendix ii). Each marker was initially sequenced in both the forward and reverse directions using the mapping family parents. If an informative SNP was discovered then the marker would be sequenced in either the forward or reverse direction depending to sequence the marker.

Species	Marker	Chr (B. mori)	Forward primer (5'-3')	Reverse primer (5'-3')	Ta C	Extens -ion time (sec)	F/R
P. pilosaria	Sulfamidase	17	GCAGACCGAGGAGAACAATCA	GTACGGAGGCTGCGTGTG	57	60	F
P. pilosaria	rab-ggt β	17	CAACTTTGAAAACTAAGCGCTGTTC	AAGCTGCCGTCGTCTTGCTG	57	30	R
P. pilosaria	shuttlecraft	17	GGAACATCTGCGCCAAGCC	GAGGTCGCCGCAGGGTATCTC	57	120	F
P. pilosaria	eIF3D	17	TGGACAAGCGCGACAACAC	AACTTGTACTTGGGCTCGGTCGT	57	90	F
P. pilosaria	apyrase	17	CCTCCGGTGATTCTGAGTTCGAG	CCCATTGAGCCGACATATAGGAT	57	60	R
E. kühniella	Simkin	18	TCGCCTTGATGATGTACT	GTGAAAGAAGTGTTCAACAA	53	70	R
E. kühniella	136	24	TACACAGGGATCTCAAGG	ATGTTCATCTCGCCGTC	43	120	F
E. kühniella	GAPDH	3	ATGCCCCCATGTTTGTTGTCGGT	CGGTATATCCTAAGACGCCCTTC	57	60	R
E. kühniella	EF1	17	GGAACGACTTGCTGCATACAATGCT	CATAAGCTAATGGCACAAGTTTGGA	57	90	F
E. kühniella	Fizzy	17	CCATAAGTGGAGCTGGCGCAC	CACCATCTTGCTGCGGACGTTGT		90	R
E. kühniella	wingless	4	TGGGCGACGCTTTGAAAGATC	ACCACGAACATGGTCTCCGT	57	90	F
E. kühniella	Attractin	26	GAACCTCCAGCAGTTCTTCA	CGACGCCATTTGTTCCATTT	53	120	R
Degenerate:							
P. pilosaria	rab-ggt β	17	NA	NA	57	60	F&R
P. pilosaria	Fizzy2 CTY	17	CTAGYTTRGTGCCRTCTGGACTGAARACCATRGT	GGAAGGAAGTTCCAGATDATRAGRTCYTCRTC	57	30	F&R
P. pilosaria	Fizzy2 TTR	17	CTAGYTTRGTGCCRTCTGGACTGAARACCATRGT	GGAAGGAAGTTCCAGATDATYAARTCYTCRTC	57	30	F&R

Table 5.1: Primer and optimal PCR amplification conditions for markers used in this study. Marker amplification information including primer pairs, species used in, orthologous chromosome that marker lies on in *B. mori*, optimal annealing temperature (TaC), optimal extension time and whether the marker was sequenced with the forward or reverse primer.

Data analysis

Once sequenced, the forward and reverse sequences for each marker were assembled using Geneious Pro v.6.0.3 (Biomatters). Sequence generated from the *P. pilosaria* mapping family was scrutinised for SNPs that were heterozygous in the male parent and homozygous in the female parent. Sequence generated from the *E. kühniella* female-informative mapping families was scrutinised for SNPs that were heterozygous in the female parent and homozygous in the male parent. Sequenced offspring from the mapping families was scored as being either homozygous or heterozygous at each marker.

In the case of the female-informative mapping families, created for *E. kühniella*, a linkage group would be identified as being associated with melanism if the informative marker co-segregates with all melanic offspring, but not with any of the non-melanic offspring.

The genomic region in which the locus that controls the *monacharia* phenotype in *P. pilosaria* is located can be delimited by assessing the segregation pattern of the paternal allele in the offspring of the *P. pilosaria* mapping family. Co-segregation between the paternal allele and melanic individuals would indicate strong linkage between that marker and the locus that controls melanism. Co-segregation was assessed by eye and confirmed with a two-tailed Fisher's exact test (tables 5.2, 5.3).

A linkage map of the *P. pilosaria* region orthologous to *B. betularia* chromosome 17, where the carbonaria locus is located, was created using Joinmap v4.1 (Stam 1993) and a map created using MapChart v2.3 (Voorrips 2002).

Results

Phigalia pilosaria

The region that controls melanism in *P. pilosaria* was mapped to a region syntenic to that which controls melanism in *B. betularia* (figure 2). The markers Rab-ggt β and shuttlecraft were both found to be associated to the linkage group that controls melanism in *P. pilosaria* (Fisher's exact test p<0.001; table 5.2). These markers are 7.1cM apart and frame the *monacharia* region, as a single recombinant was found for different individuals at each of these markers. The other three markers (sulfamidase, eIF3D and apyrase), which are located in LG 17 in *B. Betularia*, showed no association with melanism and were all found to reside in different linkage groups in *P. pilosaria* (figure 5.3).

Locus	Maternal Allele		Paternal Allele			
	monacharia	***typica***	***monacharia***	***typica***	**n**	**Fisher's Exact Test**
Sulfamidase	12	6	5	9	32	p>0.05
rab-ggt β	15	1	0	14	30	p<0.001
shuttlecraft	1	15	16	0	32	p<0.001
eIF3D	9	10	6	4	29	p>0.05
apyrase	8	6	9	9	32	p>0.05

Table 5.2: Test of association for *monacharia* and *typica* forms of *P. pilosaria* at five markers. Segregation of maternal and paternal alleles amongst F_2 offspring in *P. pilosaria* mapping family showing segregation between the phenotypes of F_2 offspring and whether their genotype matches the maternal or paternal (male-informative) allele. Significance of association was determined using Fisher's exact test.

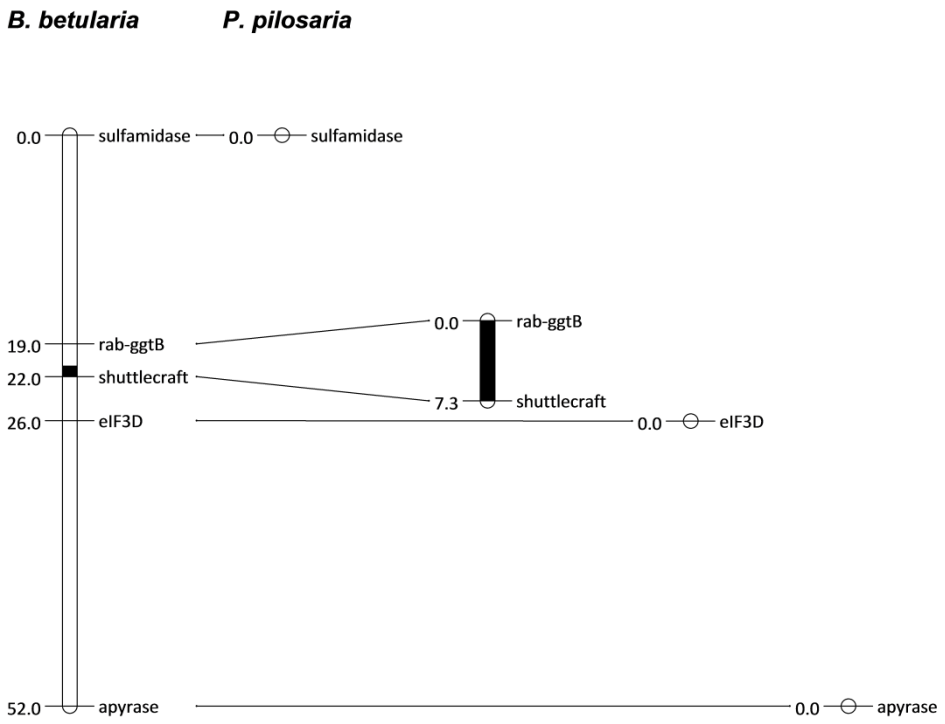

Figure 5.3: *Biston betularia* and *Phigalia pilosaria* linkage maps. Linkage map of *B. betularia* chromosome 17 (Marker information derived from (Van't Hof et al. 2013)) with the region containing the *carbonaria* locus indicated by a solid black bar. Linkage map for *P. pilosaria* using markers syntenic to those found on *B. betularia* chromosome 17. The genomic region identified as containing the *monacharia* locus is indicated by a solid black bar. Distances are indicated in cM.

Ephestia kühniella

There was no association found between melanic phenotypes in the three melanic strains of *E. kühniella* and the Lepidoptera wing patterning hotspot on *Bm* chromosome 17 (table 5.3).

Further linkage groups were analysed as marker variability permitted. Melanism in the *Ala nigra* strain was not found to be associated with markers residing in linkage groups 3 (contains no known genes in insect melanisation pathway), 18 (*arylalkylamine N-acetyltransferase)* and 24 (Ferritin heavy chain, Ferritin light chain). No association between melanism in either linkage group 3 (no known melanisation genes) or in linkage group 26 (NBAD synthase) in the *mel* strain. There was also no association between melanism and linkage groups 4 (dopa decarboxylase) and 26 (NBAD synthase) in the *abt* strain. There was no association between sex and phenotype in any of the *E. kühniella* strains, so the sex

chromosomes, which are known to contain the melanisation pathway genes phenylalanine hydroxylase, NBAD hydrolase and tyrosine hydroxylase, can also be ruled out of containing genes that control melanic morphs in any of the 3 strains.

			Female Allele		Male Allele		
Strain	**Marker**	**Chromosome (B. mori)**	**Melanic**	**Typical**	**Melanic**	**Typical**	**n**
Ala nigra	**GAPDH**	**3**	4	2	2	3	11
Ala nigra	**Simkin**	**18**	1	3	5	3	12
Ala nigra	**Ek136**	**24**	2	3	2	3	10
Ala nigra	**Fizzy**	**17**	1	1	1	1	4
abt	**wingless**	**4**	6	4	2	4	16
abt	**Attractin**	**26**	1	0	1	2	4
abt	**EF1**	**17**	4	5	3	3	15
mel	**GAPDH**	**3**	3	0	2	0	5
mel	**Attractin**	**26**	2	3	3	2	10
mel	**Fizzy**	**17**	3	9	9	1	22

Table 5.3: Test of association in three melanic strains of E. *kühniella*. Segregation of maternal and paternal alleles amongst offspring in *E. kühniella* female-informative mapping families for three melanic strains of *E. kühniella* (*Ala nigra, abt, mel*). The table details which markers were used in each *E. kühniella* strain, which chromosome each marker is located in *B. mori*. The presence of a single recombinant between allele and phenotype (melanic or typical) indicates that that linkage group is not associated with the control of melanic phenotypes.

Discussion

The genetic basis of wing patterning in Lepidoptera is an emerging model for the study of convergent evolution across numerous levels from molecule to phenotype. In Lepidoptera a hotspot for evolutionary relevant mutations involved in wing patterning has been identified on chromosome 17. Mutations in this region have been found to control wing pattern variation in *Heliconius* species, eye spots in *B. anynana,* melanism in *B. betularia* and a strain of *B. mori* (Ito *et al*., 2016). This chapter examines the role of this wing patterning hotspot in the control of melanic morphs of three Lepidoptera species: *P. pilosaria* and *E. kühniella*. The target genomic region was associated with melanism in *P. pilosaria* but not with melanic strains of *E. kühniella*.

The *monacharia* locus was successfully mapped to a region of the *P. pilosaria* genome that is syntenic to the region that control melanic forms of *B. betularia* and *B. mori,* and is also known to control aspects of wing patterning in other Lepidoptera. The 5 markers used to create the *P. pilosaria* linkage were designed to spread across a genomic region syntenic to chromosome 17 in *B. betularia* and *B. mori*, but formed 4 separate linkage groups in *P. pilosaria*. Although this may result in part due to a breakdown of linkage between distal markers, it is concordant with (Regnart 1933)'s finding that *P. pilosaria* has 112 chromosomes.

The biochemical pathway leading to melanic phenotypes is complex, and Body and wing pattern variation in the Lepidoptera can thus potentially be affected by many factors. In *B. mori* alone, over 100 mutations spread over 29 of the 31 *B. mori* chromosomes have been identified as being associated with colour and pattern variations (Dai et al. 2010; Lu et al. 2003). Control of such variation is not limited to genetic variants but also to developmental and epigenetic changes, for instance expression of the *B. anynana* big eye phenotype is dependent on whether it is the wet or dry season. Inspite of this the wide range of possible variants that control melanism, the loci that control 'natural' elements of wing patterning, such as melanism in *P. pilosaria, B. betularia*, and *B. mori* have been repeatedly found to be located across Lepidoptera families in what appears to be a 'hotspot' for evolutionarily-relevant mutations.

Patterning genes often form a part of complicated, developmental networks that has the capacity for phenotype-affecting mutations to occur anywhere within it. So why then are so many mutations that control adaptive wing pattern variation found in a subset of genes within these pathways? It is thought that such hotspots arise, at least in part, due to constraints that arise as a result of pleiotropic effects. If mutations occur early on within a pathway then they could have large knock-on effects further downstream that are negatively selected against mutations that occur towards the end of a developmental pathway and so may have a smaller effect size. There is a 'sweet spot' in these pathways where a mutation can occur that affects phenotype without having negative fitness effects. Pigmentation, particularly melanin, is involved in many adaptive traits, including insect immunity (Nappi & Christensen 2005). Changes involving pigmentation could likely have negative pleiotropic effects, causing evolutionary constraint. Evolutionary diversity may be more likely to be generated from changes occurring in *cis*-regulatory regions, which have fewer pleiotropic effects than changes to coding regions (Stern & Orgogozo 2008; Wittkopp & Beldade 2009). Genes may also be co-opted for adaptation more commonly than others if they have higher mutation rates or higher levels of standing genetic variation.

Melanic forms of *P. pilosaria* and *B. betularia*, but not *E. kühniella*, are controlled by changes to the same genomic region. *P. pilosaria* and *B. betularia* belong to the Geometridae family of moths, whilst *E. kühniella* is a member of the more distantly-related family Pyralidae. Convergent evolution is more likely to occur in closely related species (Conte et al. 2012) either as a result of shared standing genetic variation (Burke et al. 2014; Martin & Orgogozo 2013; Teotonio et al. 2009) or due to reversion to a shared, ancestral state (Martin & Orgogozo 2013). Rapid adaptation is more likely to result from standing genetic variation than *de novo* mutation (Barrett & Schluter 2008). The observation that evolution of industrial melanism in both *P. pilosaria* and *B. betularia* occurred rapidly and melanic forms of *P. pilosaria* existed at low frequencies prior to the selective event, indicates that convergence of this trait may have come about as a result of shared standing genetic variation. However, this doesn't explain convergence at this locus across more distantly related Lepidoptera, which is suggestive of a deeply conserved genetic mechanism for adaptive variation of wing patterning.

The melanic strains of *E. kühniella* used in this study arose from laboratory-maintained stocks and are not known to occur in the wild. It is possible that these melanic mutants have negative fitness costs and only exist because they are maintained in an artificial environment. The *Ala nigra* strains seems to be similar to the wild-type in terms of viability (Cotter 1974) but this doesn't exclude negative fitness effects that may arise in the wild. Hence, whilst uncovering the loci that control these melanic morphs would be informative in terms of establishing genetic changes from which melanism is able to arise, it would not necessarily be informative in terms of establishing loci from which melanic phenotypes that are positively selected for arise. Changes that give rise to the latter may well be constrained to 'hotspots' within the Lepidopteran, and hotspots thus exist because these are the loci in which pleiotropic effects of the mutation are limited. Comparison of these two sorts of loci would aid understanding of why hotspots occur in some regions of the genome and not others.

Future work should determine and understand the exact mechanisms through which melanic morphs arise in Lepidoptera. This could be achieved in *P. pilosaria* by conducting further recombinational analysis with a larger mapping family. This approach however may not be productive, as a much larger family would need to be created to increase the chance of informative recombination events in the offspring and it is difficult to raise large numbers of larvae through to pupation as they are susceptible to disease. It would also be a time-consuming process as *P. pilosaria* is univoltine. Alternatively, a population genomics approach could be adapted whereby melanic and non-melanic individuals' genomes are sequenced and aligned to a high-quality reference assembly for the *monacharia* region. Variants associated with melanism could then be assessed to determine whether they are the causative variant.

The loci that control melanism in the three melanic strains of *E. kühniella* were not identified within this study. Recombinational mapping, one marker at a time is a laborious process, compared with NGS methods that have arisen to perform the same function. To this end, RAD-seq could be used to create a genome-wide linkage map with the current mapping families. Coupled with a genome assembly, this could be used to identify genetic variants associated with melanism. However, this may still be challenging. The melanic strains used in this study were all derived from inbred laboratory stocks. Compared to

mapping families created with wild-caught individuals there were far fewer informative SNPs that could be used for mapping in the *E. kühniella* strains. Not only were there fewer polymorphisms between the parents but there were also indel differences between the parents of each strain that could not be resolved in the offspring through Sanger sequencing. This restricted the ability to map traits. This would be less of an issue if an NGS approach was taken, as there would be much more available sequence from which informative SNPs could be drawn. Another approach to alleviate this issue would be to create mapping families by crossing the melanic strains to a more distinct strain of *E. kühniella*.

Once the exact mutations that control melanism in *P. pilosaria* and *E. kühniella* have been identified, the mechanisms through which they operate can be investigated. A transcriptomics approach, coupled with knowledge of the pathways in which the melanism-controlling genes operate, would increase understanding of the processes and mechanisms through which melanism evolves and its effect on upstream and downstream processes. This would help to establish the evolutionary process and mechanisms that underpin this apparent hotspot for lepidopteran wing patterning. Similar studies are also required in other Lepidoptera to establish the extent of a shared, underlying mechanism.

This chapter investigated to what extent the phenotypic convergent evolution of melanism in Lepidoptera is accompanied by convergence at more basal levels. Specifically, this chapter investigated whether melanism in the melanic *monacharia* form of *P. pilosaria* and in three melanic strains of *E. kühniella* is controlled by a region of chromosome 17 that appears to be a hotspot for evolutionarily-relevant mutations in lepidopteran wing patterning. This region was indeed implicated in the control of melanism in *P. pilosaria*, but was not found to control melanism in *E. kühniella*. The loci that control melanism in *E*. *kühniella* remain unclear. These findings add weight to the concept that evolutionarily-relevant mutations are not found stochastically across the genome but may, at least to a certain extent, be constrained and that this constraint can give rise to 'hotspots'. The lepidopteran chromosome 17 wing patterning region appears to be one such hotspot. However, this hotspot region has neither been implicated in the control of melanic strains of *E. kühniella*, nor in the control of melanic phenotypes of a number of other Lepidoptera (*M. sexta, P. glaucus, S. exigua*). Rather, the molecular changes involved in the control of these

phenotypes appear to arise within parts of the melanin synthesis pathway, which may be another example of evolutionary constraint. Developmental pathways are complex; there may well be multiple, but not infinite, ways in which these pathways can be altered to make a phenotype. The ever expanding generation of genomic data will no doubt help to elucidate the extent to which evolution is constrained and how and why this constraint comes about.

Chapter 6

General Discussion

Identifying the loci responsible for evolutionary change falls firmly into the 'stamp collecting' category of science; the usefulness and insight gained from such endeavours has recently been called into question (Lee et al. 2014; Rausher & Delph 2015; Rockman 2012; Travisano & Shaw 2013). There does however, appear to be an emerging consensus as to the circumstances in which identifying evolutionary-relevant mutations is scientifically worthwhile (Lee et al. 2014; Rausher & Delph 2015). This thesis worked towards establishing the evolutionary-relevant mutations in two study systems that fall within this remit, namely suppression of *Wolbachia*-induced male-killing in *H. bolina* and industrial melanism in Lepidoptera. The adaptive and ecological value of these traits has been well-studied, so uncovering the genetic basis of these traits would produce case studies where we have an integrated knowledge of the processes of evolution that spans from a population to an individual level and from the phenotypic to the genotypic level. Furthermore, these case studies also contribute to the ongoing debate in evolutionary biology concerning parallel evolution, and the extent to which parallel phenotypic evolution results from convergent molecular evolution.

This thesis investigated the loci responsible for rapid evolutionary change in two lepidopteran study systems:

1. *Hypolimnas bolina/Wolbachia* system: Some *H. bolina* populations are infected with a male-killing strain of *Wolbachia*, which results in female-biased sex ratios. *H. bolina* has evolved the ability to suppress *Wolbachia*'s male-killing phenotype, so that infected females are able to produce viable male and female offspring. Selection for the suppressor is very strong.

2. Evolution of melanism in Lepidoptera: Melanism has evolved multiple times in many Lepidopteran species. Industrial melanism evolved across a number of moth species, most famously in the peppered moth, during the industrial revolution. Selection for melanic morphs was very strong, as they were much more cryptic than their typical counterparts in a polluted, urban environment. Non-industrial melanism is also found amongst the

Lepidoptera. Specifically, two species were studied within this system:

a. Pale brindled beauty *Phigalia pilosaria* - a moth with industrial melanism

b. Mediterranean flour moth *Ephestia kühniella* - a laboratory model with melanic strains

The aims of the thesis were as follows:

1. Establish whether suppression of male-killing in *H. bolina* is a single or multi-locus trait.

2. Narrow down the genomic region in which the suppressor is located, thus reducing the number of candidate genes.

3. Investigate *doublesex* as a candidate gene for suppression.

4. Explore the dynamics involved in a selective sweep for the suppressor across both space and time.

5. Investigate whether melanism in *P. pilosaria* and *E. kueniella* is controlled by a genomic region that is thought to be an evolutionary hotspot for wing patterning in Lepidoptera.

The results of this investigation are summarised below.

In chapter 2 I investigated whether a genomic region, previously identified as containing a locus necessary for male survival in suppressor present individuals, is the sole region necessary for suppression. I found no evidence for the existence of a secondary suppressor locus on any other autosome or on the sex chromosomes. This confirms that the suppression of *Wolbachia*-induced male-killing in *H. bolina* lies within a single genomic region on chromosome 25.

In Chapter 3 I further investigated the genomic region containing the suppressor. I narrowed down this region from a range spanning 10cM to one which spans only 2.4cM. This reduced the number of candidate genes within the region. The *doublesex* gene is one such candidate. Based on its role in insect sex determination and altered expression in male-killed *O. scapulalis* (Sugimoto et al. 2010; Sugimoto & Ishikawa 2012). I then went on to investigate *doublesex* as a candidate suppressor gene and found that *doublesex* lies within a region of *H. bolina* chromosomes 25 (the chromosome on which the suppressor is located) with the greatest signature of selection, associated with a selective sweep for the

suppressor on Upolu, Independent Samoa. I compared a marker for *doublesex* in *H. bolina* populations, in which the suppressor is present, with populations in which the suppressor is absent. I noticed that almost all of the individuals derived from populations in which the suppressor is present were heterozygous at this marker. I hypothesise that this apparent excess of heterozygotes in populations with the suppressor results from individuals in these populations carrying two copies of *doublesex*. From this I develop the hypothesis that the suppression of *Wolbachia*-induced male-killing in *H. bolina* is associated with duplication of *doublesex*. The strong selection associated with suppression, compared to the long stasis in which suppression did not evolve, supports involvement of a 'rare mutation'. The unusual nature of the mutation is also supported by evidence of substantial novel genetic material associated with spread of the suppressor in Independent Samoa. This indicates that the suppressor arose in Independent Samoa through immigration rather than *de novo* mutation, supporting constraint (chapter 3, 4).

In chapter 4 I looked at the dynamics of a selective sweep for the suppressor which occurred in Independent Samoa. Upolu and Savai'i are the two main islands of Independent Samoa. *H. bolina* on these islands exhibited sex ratios that were extremely female-biased as a result of *Wobachia*-induced male-killing. A 100:1 female-biased sex ratio had existed on Independent Samoa for at least one hundred years up until 2005 when it was noted that males were seen more frequently on Upolu but not on Savai'i (Charlat et al. 2007b; Dyson & Hurst 2004). A rapid selective sweep for the suppressor occurred. By 2006 the suppressor was ubiquitous on Upolu but only present on Savai'i in areas neighbouring Upolu. By 2007 the suppressor had spread across both Upolu and Savai'i (Charlat et al. 2007b). Using samples collected from before, during and after this selective event I analysed the selective sweep for the suppressor across both space and time. The selective sweep on Savai'i has the signature of a rapid and intense selection. The selective sweep on Savai'i was weaker than that which occurred on Upolu. This is consistent with theory which predicts that selective sweeps become weaker as they travel across both time and space, as there has been a longer period of time for recombination to break down associations between the selected allele and hitchhiking variants (Barton 2000; Kim & Maruki 2011). Analysis of the Savai'i sweep also illustrated how selective sweeps are able to alter the level of genetic differentiation between geographically distant populations.

In chapter 5 I moved on to study the loci of adaptation in a different study system, namely melanism in Lepidoptera. Convergent evolution of melanism has occurred throughout the Lepidoptera. Notably, there was strong selection for melanic morphs during the industrial revolution (Majerus 1998). One question that is currently of interest in evolutionary biology is the extent to which the convergent evolution of phenotypes, such as melanism, results from convergent evolution at a molecular level (Stern & Orgogozo 2008; Stern & Orgogozo 2009). Are there genomic 'hotspots' responsible for convergent phenotypes? A potential hotspot for wing patterning in Lepidoptera has been identified (Beldade et al. 2009; Ito et al. 2016; Joron et al. 2006; Kronforst et al. 2006; Papa et al. 2008; van't Hof et al. 2011). Industrial melanism in the peppered moth has been found to be controlled by this hotspot region (van't Hof et al. 2011). In this chapter I investigated whether this region is also implicated in the evolution of industrial melanism in *P. pilosaria* and in melanic strains of *E. kühniella*. I discovered that melanism in *P. pilosaria* was indeed controlled by a genomic region syntenic to the Lepidoptera wing patterning hotspot but that this was not the case for the melanic strains of *E. kühniella*. This finding adds further support to the notion that this region is an evolutionary hotspot for wing patterning in Lepidoptera but also that phenotypic convergence doesn't necessarily result from convergence at more basal levels.

The hypothesis that the suppression of *Wolbachia*-induced male-killing in *H. bolina* is associated with duplication of *doublesex* is perhaps the most interesting finding to emerge from this thesis. If suppression of male-killing does result from duplication of *doublesex* then it is an interesting finding for a number of reasons. It was once thought that *doublesex* was merely a sex differentiation gene. However, there is a growing body of evidence showing the co-option of *doublesex* for a number of sex-specific adaptations and as such, *doublesex* could be considered to be a hotspot for adaptation. Insect sex determination systems are diverse. That *H. bolina* countered parasite-mediated male-killing with changes in its sex determination system raises the question of whether parasites have driven adaptive diversification of insect sex determination systems more generally.

Within the insect sex determination system *doublesex* plays a key role signalling whether cells should develop male or female characteristics. *doublesex* has also been identified as a locus that regulates sex-specific secondary sex characteristics more generally, examples of which include: female mimetic forms of *Papilio polytes* (Kunte et al. 2014), horns in the

beetles *Onthophagus taurus* (Kijimoto et al. 2012) and *Trypoxylus dichotomus* (Ito et al. 2013), sex combs in *Drosophila melanogaster* (Devi et al. 2013; Tanaka et al. 2011), mandible growth in the stag beetle *Cyclommatus metallifer* (Gotoh et al. 2014) and social behaviour in *D. melanogaster* (Siwicki & Kravitz 2009). This evidence suggests that *doublesex*, rather than being a conserved master regulator locus, is additionally a hotspot for the evolution of adaptive, sexually dimorphic traits.

The loci of adaptation for two study species, namely suppression of *Wolbachia*-induced male-killing in *H.bolina* and the evolution of melanism in *P. pilosaria*, have found to reside in genomic regions that appear to be hotspots for the evolution of insect sexual dimorphism and lepidopteran wing patterning, respectively. The regulatory networks and molecular mechanisms in which the wing patterning hotspot operates are uncharacterised. Contrastingly, we have quite a good knowledge of the workings of the insect sex determination and sex differentiation pathways that *doublesex* is involved in. This knowledge may enlighten us as to why *doublesex* is so amenable to evolutionary change and makes it a hotspot for adaptation.

Within insect sex determination *doublesex* acts as a 'nexus' between sex determination and sexual differentiation. When required it provides sex-specific information to tissues for sexual differentiation (Verhulst & van de Zande 2015). There are many downstream targets of DSX (Clough et al. 2014; Luo et al. 2011). Its pleiotropic effects mean that although it is subject to positive selection, it has largely evolved under selective constraint (Cande et al. 2014; Eirin-Lopez & Sanchez 2015; Price et al. 2015; Sobrinho & de Brito 2012). *doublesex* pleiotropy also constrains the evolution of downstream targets, which have conserved DSX binding sites (Luo & Baker 2015). Changes in *cis*-regulatory regions have less of a pleiotropic effect than coding region changes and so drive diversity (Gompel et al. 2005; Prud'homme et al. 2006; Stern & Orgogozo 2008). Variation is commonly found to result from changes in expression of *doublesex* (Tanaka et al. 2011) and changes in the *cis*-regulatory regions of target genes (Luo & Baker 2015; Shirangi et al. 2009; Williams et al. 2008). Alternative splicing of *doublesex* also creates a diversity of (female) DSX isoforms (Duan et al. 2013b; Wang et al. 2014).

Although *doublesex* is under constraint, it still experiences positive selection. Sex is a very powerful driver of evolution (Darwin 1871), and changes in *doublesex* that can give a sex-specific advantage may be able to overcome negative effects of pleiotropy. The central role that *doublesex* inhabits within the sex determination cascade, means that it can both produce global changes within an organism, which may be subject to strong selection, but can also act locally and target specific tissues, which can reduce pleiotropic interference (Luo & Baker 2015; Verhulst & van de Zande 2015). Thus, *doublesex* could be a hotspot because changes to it can have global effects on sex-specific features, which are very strongly selected for.

Endosymbionts that manipulate host sex ratio are known to have different targets within the sex determination system. *Wolbachia* in *O. scapulalis* and *O. furnacalis*, and *Spiroplasma poulsonii* in *D. melanogaster*, cause male-killing in their respective hosts by inhibiting Z/X chromosome dosage compensation (Cheng et al. 2016; Fukui et al. 2015; Sugimoto et al. 2015; Veneti et al. 2005). In *O. scapulalis, Wolbachia* is also known to interfere with *doublesex* expression with genetic males expressing female forms of *doublesex*, which leads to male death (Sugimoto & Ishikawa 2012). In *D. melanogaster*, *S. poulsonii* male-killing is associated with the misdirection of the dosage compensation complex to the autosomes and in this case the interaction occurs directly with the dosage compensation system (Cheng et al. 2016; Veneti et al. 2005). *Arsenophonus* is known to cause male-killing in *Nasonia* by interfering with the formation of maternal centrosomes (Ferree et al. 2008). The woodlouse *Armidillidium vulgare* carries feminising strains of *Wolbachia* (Bouchon et al. 1998; Cordaux et al. 2004; Rigaud et al. 1997; Rigaud et al. 1999). *Wolbachia* infection causes ZZ individuals to develop as females. In infected populations the female W chromosome has been lost and sex is determined by either the presence (female) or absence (male) of *Wolbachia*. A masculinizing gene (*M*) has evolved that is able to fully suppress the feminising effects of a feminising *Wolbachia* nuclear insertion, and promotes the development of intersex females that are capable of transmitting the M gene in *Wolbachia* positive individuals (Rigaud & Juchault 1993).

The diversity of targets of male-killing bacteria suggests that it will not simply be *doublesex* that is the target of selection to rescue male function. There is thus an expectation that the evolution of suppression where it occurs may drive the evolution of multiple elements of

the sex determination cascade. These cases would add to those in which *Wolbachia* completely hijack its hosts' sex determination functions and becomes required for the normal sex determination of female *O. scapulalis* and *Eurema mandarina* (Kern et al. 2015; Sugimoto & Ishikawa 2012). Furthermore, female *E. mandarina* now only inherit a single Z chromosome (Kern et al. 2015).

Thus, endosymbionts are able to cause changes to their hosts' sex determination system, but there is strong selection on the host to counteract the changes inflicted upon it by its endosymbiont. This has occurred in *H. bolina*, which has evolved the ability to suppress male-killing (Hornett et al. 2006). If, as hypothesised, suppression results from duplication of *doublesex* then this further demonstrates that endosymbionts act as drivers for the adaptive diversification of sex determination system evolution. The array of sex determination systems in insects could have evolved as a result of endosymbiont-driven adaptation, making endosymbionts drivers of core developmental pathways (Beukeboom 2012; Cordaux et al. 2011). A key question here is how commonly does evolution of suppression of male-killing evolve? For *Wolbachi*a, it is possible that the evolution of suppression is common, but rarely observed because male-killing becomes a hidden phenotype, with the suppressed strain exhibiting CI. This is the case in *H. bolina* post suppressor spread, and also appears to have occurred in the moth *Ephestia cautella*. In this host species, the *Wolbachia* causes CI, but exhibits male-killing following transinfection to *E. kühniella*, compatible with it being a suppressed male-killer in the native host (Jaenike 2007b). However, there are also many cases where *Wolbachia* transinfection has been conducted without the emergence of male-killing.

For non-*Wolbachia* male-killers (where CI is presumed not to occur), suppression is expected to be associated with reduced frequency or loss of male-killing (Randerson et al. 2000). The ladybird *Cheilomenes sexmaculata* carries both a male-killer and a polymorphic suppressor (Majerus & Majerus 2010a), and this co-existence is expected where suppression is costly. Cost free suppressors, however, will lead to the loss of the symbiont unless it has alternate forms of drive (such as a direct benefit); at this point, suppressor evolution will be cryptic.

Outstanding questions still remain in both of the study systems examined during the course of this thesis. In the *H.bolina*/*Wolbachia* system the key question to be asked now is whether suppression is caused by duplication of *doublesex*, if so how does this change enable male survival? Fitness consequences associated with suppression should also be studied – is the fitness of males carrying the suppressor different to uninfected males without the suppressor, does the suppressor affect female fitness, if so in what ways is fitness affected? Another 'open question' is to understand exactly how *Wolbachia* is able to manipulate its hosts' sex determination system: whilst the host systems that Wolbachia interferes with are becoming clear in an increasing number of cases, the effectors that cause the interference remain obscure.

In the melanism in Lepidoptera study system the key questions that remain unanswered include: what exactly are the loci that control melanism in *P. pilosaria* and in *E. kühniella*. What makes the wing patterning region on chromosome 17 special, why are the loci of so many adaptive wing patterning traits located here? Do similar changes within this region result in similar phenotypes, even in distantly related Lepidoptera species? How often do adaptive changes occur within this region in Lepidoptera more broadly?

To conclude, in this thesis I investigated the evolutionary genetics in two lepidopteran study systems in which rapid adaptation has occurred: The evolution of the suppression of *Wolbachia*-induced male-killing in *H. bolina*, and the evolution of melanism in Lepidoptera. In the former system I confirmed that suppression of male-killing is a single locus trait, proposed the hypothesis that the change that occurred at this locus was the duplication of *doublesex*, and further characterised a selective sweep for the suppressor in Independent Samoa across both time and space. In the latter study system I showed that the locus that controls industrial melanism in *P. pilosaria* is syntenic to loci found to control adaptive wing patterning traits in other Lepidoptera species, but found no evidence that this locus was responsible for controlling melanism in *E. kühniella* strains. These findings have wider implications in terms of informing our knowledge of the processes of adaptation, whether endosymbionts act as drivers of developmental systems, and to what extent evolutionary 'hotspots' are involved in adaptation.

Appendix

i. DNA extraction protocol using Phenol/Chloroform

- Homogenise a piece of tissue approximately 5mm^3 in size with 500µl insect extraction buffer (50mM Tris HCl pH 8.0, 25mM NaCl, 25mM EDTA pH 8.0, 0.1% SDS)
- Add 4 µl proteinase K (10mg/ml) to each sample
- Incubate at 55°C for 2 hours, mixing every 30 minutes
- Add 2µl RNAse A (10µg/ml) to each sample
- Incubate at 37°C for 20 minutes
- Add 500 µl phenol/chloroform to the lysed sample.
- Mix thoroughly
- Spin samples at 13 000 rpm for 5 minutes
- Add 400 µl of the supernatant to 400 µl of chloroform
- Mix thoroughly
- Centrifuge at 13 00 rpm for 1 minute
- Add 340 µl of supernatant to 680 µl 100% ice cold EtOH + 30 µl NaoAc pH 5.2 Store samples at -20°C for 2 hours
- Centrifuge at 4°C at maximum speed for 10 minutes
- Add supernatant to 750 µl 70% ice cold EtOH
- Centrifuge at 4°C at maximum speed for 2 minutes
- Remove supernatant and allow the sample to briefly air dry (do not allow to completely dry out)
- Add 200 µl ddH_20, re-suspend DNA pellet by incubating at 37°C Asses DNA for $_q$uality by running 2 µl on a gel and by analysing 1 µl using a nanodrop

ii. PCR product purification and sequencing protocol

1. Purification of PCR product using ExoSAP digest clean up (perform on ice)

- Make up a mastermix containing: 0.2 µl SAP, 0.05 µl exonuclease I, 0.7 µl 10X exonuclease I reaction buffer, 3.05 µl ddH_2O.
- Combine 4 µl of mastermix with 3 µl PCR product
- Incubate at 37°C for 45 minutes followed by 15minutes at 80°C

2. Cycle sequencing reaction

- Make up a mastermix containing: 1µl 5X BigDye 3.1 reaction buffer, 10pmol of primer, 7.5 µl ddH_2O, 0.18 µl Bigdye Terminator v3.1
- Combine 9 µl sequencing mastermix with 1 µl of purified PCR product from stage 1
- Incubate in a thermocycler for 29 cycles consisting of:
    - 96C for 10 seconds
    - 50C for 5 seconds
    - 60C for 4 minutes

3. Cycle sequencing reaction precipitation clean up

- Make up a mastermix containing: 1.54 µl 3M Sodium Acetate (pH4.6), 32.08 µl 100%EtOH, 6.38 µl ddH_2O
- For each sample, combine 10 µl of cycle sequencing reaction from stage 2 with 40 µl of precipitation clean up mastermix
- Leave at room temperature for 15 minutes
- Spin the plate at 1400-2000g (~3100rpm) for 45 minutes
- Discard supernatant by inverting the plate onto a paper towel
- Spin the inverted plate at 600rpm for 1 minute
- Add 150 µl 70% EtOH to each sample
- Spin the plate at maximum speed (3500rpm) for 10 minutes
- Discard supernatant by inverting the plate onto a paper towel
- Spin the inverted plate at 600rpm for 1 minute
- Add 10 µl HiDi Formamide to each sample
- Vortex

4. Sequence samples using ABI3130 Genetic Analyzer

iii. RNA extraction protocol

- Grind a piece of tissue approximately 5mm3 in a liquid nitrogen cooled pestle and mortar
- Add the homogenised tissue into 1000 µl TRIZOL and mix
- Incubate at room temperature for 5 minutes
- Add 200 µl chloroform
- Mix tubes thoroughly for ~15 seconds
- Incubate at room temperature for 2 minutes
- Centrifuge at 11 200 rpm for 15 minutes at 4°C
- Add the aqueous phase to 500 µl isopropanol
- Mix gently and incubate at room temperature for 10 minutes
- Centrifuge at 11 200 rpm for 10 minutes at 4°C
- Add the supernatant to 1000 µl 75% EtOH
- Vortex
- Centrifuge at 8883 rpm for 5 minutes at 4°C
- Remove the supernatant
- Briefly air-dry the pellet
- Re-suspend the RNA in 50µl RNAse-free H_2O by incubating at 55°C for 10 minutes
- Assess for quality by running 2 µl of RNA on a gel and by analysing 1 µl using a nanodrop
- Store samples at -80°C

iv. 1st strand cDNA synthesis protocol

- Add 5 µl RNA to 0.5 µl 100µM oligo(dT)20, 1 µl dNTPs, 7.5 µl ddH_2O
- Incubate at 65°C for 5 minutes and then on ice for 1 minute
- Briefly centrifuge
- To the sample add: 4 µl 5 x 1st strand buffer, 1 µl 0.1M DTT, 1 µl SuperScript III RT
- Mix by gentle pipetting
- Incubate at 50°C for 60 minutes followed by 70°C for 15 minutes
- Add 1 µl RNAse to the sample and incubate at 37°C for 30 minutes

v. *Ephestia kühniella* mapping family crosses

♀ WT-C/WT-C x ♂An/An

Parental cross

♀ F_1 An/WT-C x ♂WT-C/WT-C

Female-informative family backcross

♀WT-C/WT-C x ♂F_1 An/WT-C

Male-informative family backcross

Figure 1: Crosses used for linkage mapping of the *An* melanic mutation in *Ephestia kuehniella.*

♀ WT-C/WT-C x ♂*abt/abt*

Parental cross

♀ F_1 *abt*/WT-C x ♂*abt/abt*

Female-informative family backcross

♀*abt/abt* x ♂F_1 *abt*/WT-C

Male-informative family backcross

Figure 2: Crosses used for linkage mapping of the *abt* melanic mutation in *Ephestia kuehniella.*

♀ wa/wa x ♂mel/mel

Parental cross

♀ F_1 mel/wa x ♂mel/mel

Female-informative family backcross

♀mel/mel x ♂F_1 mel/wa

Male-informative family backcross

Figure 3: Crosses used for linkage mapping of the *mel* melanic mutation in *Ephestia kuehniella.*

vi. Allele frequencies for *H. bolina* samples collected from Upolu in 2001, 2006, 2010, and from Savai'i in 2001, 2007, 2010.

The primary allele linked to suppression at each marker is highlighted in red, secondary alleles are highlighted in a lighter shade of red. Data for markers Hb80, Hb89, Hb90l, Hb99, Hb131, Hb136, Hb147, Hb152, Hb175, Hb176 for samples from Upolu has previously been published in Hornett *et al.*, (2014).

Hb80		Upolu #			Savai'i #			Upolu Frequency			Savai'i Frequency		
Haplotype ID	Haplotype	2001	2006	2010	2001	2007	2010	2001	2006	2010	2001	2007	2010
4	ACT	0	17	15	0	10	10	0.00	0.19	0.18	0.00	0.24	0.13
1	ATT	75	62	59	61	28	64	0.82	0.70	0.70	0.90	0.67	0.80
2	TTT	11	6	10	4	1	3	0.12	0.07	0.12	0.06	0.02	0.04
3	ATC	6	3	0	3	3	3	0.07	0.03	0.00	0.04	0.07	0.04
		92	88	84	68	42	80	1	1	1	1	1	1

Hb89		Upolu #			Savai'i #			Upolu Frequency			Savai'i Frequency		
Haplotype ID	Haplotype	2001	2006	2010	2001	2007	2010	2001	2006	2010	2001	2007	2010
1	ATTTACCC	14	52	39	9	17	35	0.15	0.58	0.50	0.14	0.40	0.44
2	ATTTAGCC	0	7	18	1	5	5	0.00	0.08	0.23	0.02	0.12	0.06
3	ATTTCCCC	17	7	4	8	5	11	0.18	0.08	0.05	0.12	0.12	0.14
4	GTGTCCTC	31	15	9	39	7	24	0.34	0.17	0.12	0.59	0.17	0.30
5	GGTCCCCC	0	1	0	0	0	0	0.00	0.01	0.00	0.00	0.00	0.00
6	GGGTCCCC	28	4	6	7	5	2	0.30	0.04	0.08	0.11	0.12	0.03
7	GGGCCCCA	2	4	2	2	1	2	0.02	0.04	0.03	0.03	0.02	0.03
8	ATTTCCT	0	0	0	0	0	1	0.00	0.00	0.00	0.00	0.00	0.01
9	GTTTACC	0	0	0	0	2	0	0.00	0.00	0.00	0.00	0.05	0.00
		92	90	78	66	42	80	1	1	1	1	1	1

Hb99		Upolu #			Savai'i #			Upolu Frequency			Savai'I Frequency		
Haplotype ID	Haplotype	2001	2006	2010	2001	2007	2010	2001	2006	2010	2001	2007	2010
1	TAAC	35	74	55	18	30	49	0.36	0.77	0.76	0.26	0.71	0.72
2	TAAT	39	7	7	24	4	6	0.41	0.07	0.10	0.34	0.10	0.09
3	TATC	14	9	7	24	7	12	0.15	0.09	0.10	0.34	0.17	0.18
4	TGAT	0	2	0	0	0	0	0.00	0.02	0.00	0.00	0.00	0.00
5	GAAT	1	4	0	4	0	0	0.01	0.04	0.00	0.06	0.00	0.00
6	TGAC	7	0	3	0	1	0	0.07	0.00	0.04	0.00	0.02	0.00
7	TGTC	0		0	0	0	1	0.00	0.00	0.00	0.00	0.00	0.01
		96	96	72	70	42	68	1	1	1	1	1	1

Hb131		Upolu #			Savai'i #			Upolu Frequency			Savai'i Frequency		
Haplotype ID	Haplotype	2001	2006	2010	2001	2007	2010	2001	2006	2010	2001	2007	2010
1	GTATC	20	42	43	6	9	12	0.21	0.45	0.47	0.09	0.21	0.15
2	GTACC	72	44	41	56	28	60	0.75	0.47	0.45	0.88	0.67	0.75
3	GTACT	2	2	0	1	1	0	0.02	0.02	0.00	0.02	0.02	0.00
4	GCACC	2	4	6	0	2	2	0.02	0.04	0.07	0.00	0.05	0.03
5	ATACC	0	2	2	0	1	5	0.00	0.02	0.02	0.00	0.02	0.06
6	GTGCC	0	0	0	1	1	1	0.00	0.00	0.00	0.02	0.02	0.01
		96	94	92	64	42	80	1	1	1	1	1	1

Hb152		Upolu #			Savai'i #			Upolu Frequency			Savai'i Frequency		
Haplotype ID	Haplotype	2001	2006	2010	2001	2007	2010	2001	2006	2010	2001	2007	2010
1	GATC	0	40	20	0	11	12	0.00	0.42	0.22	0.00	0.28	0.15
2	GTTC/GAAC	49	23	32	31	18	42	0.51	0.24	0.35	0.57	0.45	0.54
3	GTTT/AAAC	23	21	25	9	6	12	0.24	0.22	0.27	0.17	0.15	0.15
4	GTAT/ATAC	16	8	8	1	4	1	0.17	0.08	0.09	0.02	0.10	0.01
5	ATTT/AAAT	8	4	5	13	1	11	0.08	0.04	0.05	0.24	0.03	0.14
6	GATT	0	0	2	0	0	0	0.00	0.00	0.02	0.00	0.00	0.00
		96	96	92	54	40	78	1	1	1	1	1	1

Hb175		Upolu #			Savai'i #			Upolu Frequency			Savai'i Frequency		
Haplotype ID	Haplotype	2001	2006	2010	2001	2007	2010	2001	2006	2010	2001	2007	2010
1	CACCCTCCAC	20	59	45	10	19	45	0.21	0.64	0.56	0.14	0.45	0.56
2	CACCCTCCGT	6	1	2	8	6	5	0.06	0.01	0.03	0.11	0.14	0.06
3	CACCCTCTAC	4	1	0	1	0	2	0.04	0.01	0.00	0.01	0.00	0.03
4	CACCCTTCAC	11	3	1	5	0	1	0.11	0.03	0.01	0.07	0.00	0.01
5	CACTCTCCAC	0	10	15	1	6	7	0.00	0.11	0.19	0.01	0.14	0.09
6	CAACCTCCAC	46	15	16	22	8	15	0.48	0.16	0.20	0.31	0.19	0.19
7	CAACTTCCAC	0	1	0	0	0	0	0.00	0.01	0.00	0.00	0.00	0.00
8	CGCCCTCCAC	5	1	0	11	0	3	0.05	0.01	0.00	0.16	0.00	0.04
9	AGCCCTCCAC	0	1	0	0	0	0	0.00	0.01	0.00	0.00	0.00	0.00
10	CACCTTCTAC	1	0	0	0	0	0	0.01	0.00	0.00	0.00	0.00	0.00
11	AACCCTCCAC	1	0	0	0	0	1	0.01	0.00	0.00	0.00	0.00	0.01
12	AACCCCCCAC	2	0	0	0	0	0	0.02	0.00	0.00	0.00	0.00	0.00
13	AACTCTCCAC	0	0	1	0	0	0	0.00	0.00	0.01	0.00	0.00	0.00
14	CACCTTCCAC	0	0	0	9	3	1	0.00	0.00	0.00	0.13	0.07	0.01
15	CATTCTTCAC	0	0	0	3	0	0	0.00	0.00	0.00	0.04	0.00	0.00
		96	92	80	70	42	80	1	1	1	1	1	1

Hbdoublesex		Upolu #			Savai'i #			Upolu Frequency			Savai'i Frequency		
Haplotype ID	Haplotype	2001	2006	2010	2001	2007	2010	2001	2006	2010	2001	2007	2010
1	CTAA	2	0	0	0	0	0	0.03	0.00	0.00	0.00	0.00	0.00
2	CTAT	4	36	45	0	13	39	0.05	0.45	0.49	0.00	0.43	0.49
3	CCAT	21	2	1	9	0	0	0.26	0.03	0.01	0.19	0.00	0.00
4	ATCA	51	42	46	37	15	41	0.64	0.53	0.50	0.77	0.50	0.51
5	ATCT	1	0	0	1	1	0	0.01	0.00	0.00	0.02	0.03	0.00
6	ATAA	1	0	0	0	0	0	0.01	0.00	0.00	0.00	0.00	0.00
7	CTCT	0	0	0	0	1	0	0.00	0.00	0.00	0.00	0.03	0.00
8	CCCT	0	0	0	1	0	0	0.00	0.00	0.00	0.02	0.00	0.00
		80	80	92	48	30	80	1	1	1	1	1	1

HbProspero		Upolu #			Savai'i #			Upolu Frequency			Savai'i Frequency		
Haplotype ID	Haplotype	2001	2006	2010	2001	2007	2010	2001	2006	2010	2001	2007	2010
1	CTACATTTTACC	23	10	0	21	2	10	0.26	0.14	0.00	0.33	0.07	0.17
2	CTACATTTTCCC	5	0	0	9	0	2	0.06	0.00	0.00	0.14	0.00	0.03
3	CTACATTGTACC	2	1	0	0	0	0	0.02	0.01	0.00	0.00	0.00	0.00
4	CTACACTTTACC	0	1	0	1	0	0	0.00	0.01	0.00	0.02	0.00	0.00
5	CTATACTTTATC	28	5	3	3	3	0	0.32	0.07	0.03	0.05	0.11	0.00
6	CTATTCTTTATC	1	0	0	0	0	0	0.01	0.00	0.00	0.00	0.00	0.00
7	CTTCATTTTACC	5	1	0	4	0	0	0.06	0.01	0.00	0.06	0.00	0.00
8	CTTCTTTTTACC	2	0	0	0	0	0	0.02	0.00	0.00	0.00	0.00	0.00
9	CAACATTTTACT	0	50	88	0	21	46	0.00	0.71	0.96	0.00	0.75	0.77
10	CAACATATTACC	1	0	0	0	0	0	0.01	0.00	0.00	0.00	0.00	0.00
11	CAATACTTTATC	0	0	1	0	0	0	0.00	0.00	0.01	0.00	0.00	0.00
12	TTTCATTTTACC	17	2	0	1	0	1	0.19	0.03	0.00	0.02	0.00	0.02
13	TTTCATTTAACC	4	0	0	1	0	0	0.05	0.00	0.00	0.02	0.00	0.00
14	CTACTTATTACC	0	0	0	4	0	0	0.00	0.00	0.00	0.06	0.00	0.00
15	CAACATTTTCCC	0	0	0	1	0	0	0.00	0.00	0.00	0.02	0.00	0.00
16	CAACACTTTACC	0	0	0	19	2	1	0.00	0.00	0.00	0.30	0.07	0.02
		88	70	92	64	28	60	1	1	1	1	1	1

Hb5100

Haplotype ID	Haplotype	Upolu#			Upolu Frequency		
		2001	2006	2010	2001	2006	2010
1	CCAACGAT	28	59	50	0.30	0.67	0.56
2	CCAACGGT	32	9	11	0.35	0.10	0.12
3	CCAACTAT	2	2	0	0.02	0.02	0.00
4	CCAACTAA	4	0	0	0.04	0.00	0.00
5	CCAAAGAT	6	0	1	0.07	0.00	0.01
6	CCATCGAT	6	1	0	0.07	0.01	0.00
7	CCGACGAT	1	1	0	0.01	0.01	0.00
8	CGAACGAT	0	1	1	0.00	0.01	0.01
9	CGGACGAT	0	8	24	0.00	0.09	0.27
10	TCAACTAT	13	7	3	0.14	0.08	0.03
		92	88	90	1	1	1

Hb5105

Haplotype ID	Haplotype	Upolu#			Upolu Frequency		
		2001	2006	2010	2001	2006	2010
8	GGAATAATAGATGGTGGTGAA	0	38	52	0.00	0.58	0.58
5	GGAATAATAGATTGTGTTGAG	5	9	24	0.06	0.14	0.27
4	GGAATAATAGATTGTGGTGAG	37	5	10	0.45	0.08	0.11
1	GGAACAGTAGATTGTTGTTCG	26	4	3	0.32	0.06	0.03
2	GGAACAGTAGATTGTTGGTCG	2	2	0	0.02	0.03	0.00
3	GGAATAAAAGATTGTGGTGAG	8	6	0	0.10	0.09	0.00
6	GGAATAATAGATTGTGTTGAA	0	0	1	0.00	0.00	0.01
7	GGAATAATAGATGGTGGTGAG	1	1	0	0.01	0.02	0.00
9	GGTGTTATTAACTGTGGTGAG	2	1	0	0.02	0.02	0.00
10	AAAATAAAAGGTTCAGGTGAG	1	0	0	0.01	0.00	0.00
		82	66	90	1	1	1

Hb5114		Upolu#			Upolu Frequency		
Haplotype ID	Haplotype	2001	2006	2010	2001	2006	2010
1	CGGCTGATGTTA	16	2	1	0.19	0.03	0.01
2	CGGCCGATGTCA	12	2	1	0.14	0.03	0.01
3	CGGCCGATGTCT	0	0	1	0.00	0.00	0.01
4	CGGCCGATATTA	1	0	0	0.01	0.00	0.00
5	CGGCCGATACCA	10	5	1	0.12	0.06	0.01
6	CGGCCGATACCT	0	0	1	0.00	0.00	0.01
7	CGGTCGATGTCA	1	0	0	0.01	0.00	0.00
8	CGGTCGATACTA	6	0	0	0.07	0.00	0.00
9	CGGTCGATACCT	1	44	50	0.01	0.56	0.56
10	CGGTCGTTACCT	0	1	0	0.00	0.01	0.00
11	CGGTCATTACCA	0	0	1	0.00	0.00	0.01
12	CGGTCATCACCA	37	15	5	0.43	0.19	0.06
13	CAGTCGATACCT	0	1	1	0.00	0.01	0.01
14	CAACCGATGTCA	0	7	28	0.00	0.09	0.31
15	TGGCCGATGTTA	2	1	0	0.02	0.01	0.00
		86	78	90	1	1	1

Hb90

Haplotype ID	Haplotype	Upolu#			Upolu Frequency		
		2001	2006	2010	2001	2006	2010
1	ATTACGT	47	29	31	0.50	0.32	0.34
2	ATAACAC	21	7	20	0.22	0.08	0.22
3	TTTACGT	11	6	8	0.12	0.07	0.09
4	ATTGCGT	7	42	21	0.07	0.46	0.23
5	ATAATGC	4	6	5	0.04	0.07	0.06
6	AAAATGC	3	2	0	0.03	0.02	0.00
7	ATTACAC	1	0	0	0.01	0.00	0.00
8	TTTGCGT	0	0	1	0.00	0.00	0.01
9	AAAATGT	0	0	3	0.00	0.00	0.03
10	AAAACGT	0	0	1	0.00	0.00	0.01
11	ATAACGC	0	0	0	0.00	0.00	0.00
12	ATAATAC	0	0	0	0.00	0.00	0.00
13	ATAATGT	0	0	0	0.00	0.00	0.00
14	ATTATGC	0	0	0	0.00	0.00	0.00
15	ATTACAT	0	0	0	0.00	0.00	0.00
		94	92	90	1	1	1

Hb136

Haplotype ID	Haplotype	Upolu#			Upolu Frequency		
		2001	2006	2010	2001	2006	2010
1	CCG	56	78	58	0.58	0.81	0.63
2	TCG	37	17	34	0.39	0.18	0.37
3	TTA	3	1	0	0.03	0.01	0.00
4	TTG	0	0	0	0.00	0.00	0.00
		96	96	92	1	1	1

Hb147

		Upolu#			Upolu Frequency		
Haplotype ID	Haplotype	2001	2006	2010	2001	2006	2010
8	GGCTTAT	34	30	23	0.35	0.32	0.27
5	GGGCCGA	24	29	28	0.25	0.31	0.33
4	GGCTTAA	14	7	4	0.15	0.07	0.05
1	GTGTCAA	13	13	11	0.14	0.14	0.13
2	GGGACAA	6	4	4	0.06	0.04	0.05
3	GGGTCGA	3	1	0	0.03	0.01	0.00
6	GGGTCAA	2	6	8	0.02	0.06	0.09
7	CGGTCAA	0	4	8	0.00	0.04	0.09
		96	94	86	1	1	1

Hb176

		Upolu#			Upolu Frequency		
Haplotype ID	Haplotype	2001	2006	2010	2001	2006	2010
1	GTACC	0	60	55	0.00	0.65	0.69
6	CAACC	27	10	5	0.33	0.11	0.06
4	CTACC	21	9	4	0.26	0.10	0.05
3	GTATC	19	6	10	0.23	0.07	0.13
2	GTACT	8	5	3	0.10	0.05	0.04
7	GTCCT	7	0	3	0.09	0.00	0.04
5	CTCCT	0	2	0	0.00	0.02	0.00
		82	92	80	1.00	1.00	1.00

References

Arendt, J. & Reznick, D. 2008 Convergence and Parallelism Reconsidered: What Have We Learned About the Genetics of Adaptation? *Trends in Ecology & Evolution* **23,** 26-32.

Barrett, R. D. H. & Schluter, D. 2008 Adaptation from Standing Genetic Variation. *Trends in Ecology & Evolution* **23**, 38-44.

Barton, N. H. 2000 Genetic Hitchhiking. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **355**, 1553-1562.

Barton, N. H., Etheridge, A. M., Kelleher, J. & Veber, A. 2013 Genetic Hitchhiking in Spatially Extended Populations. *Theoretical Population Biology* **87**, 75-89.

Bassett, A. R., Tibbit, C., Ponting, C. P. & Liu, J. L. 2013 Highly Efficient Targeted Mutagenesis of Drosophila with the CRISPR/Cas9 System. *Cell Reports* **4**, 220-228.

Baxter, S. W., Papa, R., Chamberlain, N., Humphray, S. J., Joron, M., Morrison, C., Ffrench-Constant, R. H., McMillan, W. O. & Jiggins, C. D. 2008 Convergent Evolution in the Genetic Basis of Mullerian Mimicry in *Heliconius* Butterflies. *Genetics* **180**, 1567-1577.

Beatty, J. 2006 Replaying Life's Tape (Stephen Gould). *Journal of Philosophy* **103**, 336-362.

Beldade, P., Saenko, S. V., Pul, N. & Long, A. D. 2009 A Gene-Based Linkage Map for *Bicyclus anynana* Butterflies Allows for a Comprehensive Analysis of Synteny with the Lepidopteran Reference Genome. *Plos Genetics* **5**, e1000366.

Bentley, J. K., Veneti, Z., Heraty, J. & Hurst, G. D. D. 2007 The Pathology of Embryo Death Caused by the Male-Killing *Spiroplasma* Bacterium in *Drosophila nebulosa*. *Bmc Biology* **5**, 9.

Beukeboom, L. W. 2012 Microbial Manipulation of Host Sex Determination. *Bioessays* **34**, 484-488.

Bierne, N. 2010 The Distinctive Footprints of Local Hitchhiking in a Varied Environment and Global Hitchhiking in a Subdivided Population *Evolution* **64**, 3254-3272.

Bishop, J. A., Cook, L. M. & Muggleton, J. 1978 Response of 2 Species of Moths to Industrialization in North West England. 1. Polymorphisms for Melanism. 2. Relative Fitness of Morphs and Population Size *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **281**, 489-542.

Bishop, J. A. & Harper, P. S. 1970 Melanism in Moth *Gonodontis bidentata* - A Cline Within Merseyside Conurbation *Heredity* **25**, 449-456.

Bouchon, D., Rigaud, T. & Juchault, P. 1998 Evidence for widespread *Wolbachia* Infection in Isopod Crustaceans:Molecular Identification and Host Feminization. *Proceedings of the Royal Society B-Biological Sciences* **265**, 1081-1090.

Braverman, J. M., Hudson, R. R., Kaplan, N. L., Langley, C. H. & Stephan, W. 1995 The Hitchhiking Effect on the SIte Frequency Spectrum of DNA Polymorphisms. *Genetics* **140**, 783-796.

Bretschneider, R. 1939 Neues Uber Melanismus der Schmetterlinge im Gau Sachsen. 1. Geschlechtsbegrentzer,Dominanter Melanismus beim von *Phigalia pedaria* F. *Ent. Z. Frankf. a.M.* **53**, 59-60.

Brunet, E., Derrida, B., Mueller, A. H. & Munier, S. 2006 Noisy Traveling Waves: Effect of Selection on Genealogies. *Europhysics Letters* **76**, 1-7.

Burke, M. K., Liti, G. & Long, A. D. 2014 Standing Genetic Variation Drives Repeatable Experimental Evolution in Outcrossing Populations of *Saccharomyces cerevisiae*. *Molecular Biology and Evolution* **31**, 3228-3239.

Cande, J., Stern, D. L., Morita, T., Prud'homme, B. & Gompel, N. 2014 Looking Under the Lamp Post: Neither fruitless nor *doublesex* Has Evolved to Generate Divergent Male Courtship in *Drosophila*. *Cell Reports* **8**, 362-369.

Caspari, E. W. G., F.J. 1975 The Mediterranean Meal Moth, *Ephestia kuehniella*. *Handbook of Genetics* **3**, 125-147.

Cavalcanti, A. G. L., Falcao, D. N. & Castro, L. E. 1957 Sex-Ratio in *Drosophila prosaltans* - A Character due to Interaction Between Nuclear Genes and Cytoplasmic Factors. *American Naturalist* **91**, 327-329.

Cazemajor, M., Landre, C. & Montchamp Moreau, C. 1997 The sex-ratio trait in *Drosophila simulans*: Genetic Analysis of Distortion and Suppression. *Genetics* **147**, 635-642.

Chamberlain, N. L., Hill, R. I., Kapan, D. D., Gilbert, L. E. & Kronforst, M. R. 2009 Polymorphic Butterfly Reveals the Missing Link in Ecological Speciation. *Science* **326**, 847-850.

Charlat, S., Davies, N., Roderick, G. K. & Hurst, G. D. D. 2007a Disrupting the Timing of *Wolbachia*-Induced Male-Killing. *Biology Letters* **3**, 154-156.

Charlat, S., Engelstadter, J., Dyson, E. A., Hornett, E. A., Duplouy, A., Tortosa, P., Davies, N., Roderick, G. K., Wedell, N. & Hurst, G. D. D. 2006 Competing Selfish Genetic Elements in the Butterfly *Hypolimnas bolina*. *Current Biology* **16**, 2453-2458.

Charlat, S., Hornett, E. A., Dyson, E. A., Ho, P. P. Y., Loc, N. T., Schilthuizen, M., Davies, N., Roderick, G. K. & Hurst, G. D. D. 2005 Prevalence and Penetrance Variation of Male-Killing *Wolbachia* Across Indo-Pacific Populations of the Butterfly *Hypolimnas bolin*a. *Molecular Ecology* **14**, 3525-3530.

Charlat, S., Hornett, E. A., Fullard, J. H., Davies, N., Roderick, G. K., Wedell, N. & Hurst, G. D. D. 2007b Extraordinary Flux in Sex Ratio. *Science* **317**, 214-214.

Charlat, S., Reuter, M., Dyson, E. A., Hornett, E. A., Duplouy, A., Davies, N., Roderick, G. K., Wedell, N. & Hurst, G. D. D. 2007c Male-Killing Bacteria Trigger a Cycle of Increasing Male Fatigue and Female Promiscuity. *Current Biology* **17**, 273-277.

Chase, C. D. & Gabay-Laughnan, S. 2004 Cytoplasmic Male Sterility and Fertility Restoration by Nuclear Genes. *Molecular Biology and Biotechnology of Plant Organelles: Chloroplasts and Mitochondria*, 593-621.

Cheng, B., Kuppanda, N., Aldrich, J. C., Akbar, O. S. & Ferree, P. M. 2016 Male-Killing *Spiroplasma* Alters Behavior of the Dosage Compensation Complex during *Drosophila melanogaster* Embryogenesis. *Current Biology* **26**, 1339-1345.

Chu, Y., Zhou, F., Liu, Y., Hong, F., Wang, G. & An, C. 2015 *Ostrinia furnacalis* serpin-3 Regulates Melanization Cascade by Inhibiting a Prophenoloxidase-Activating Protease. *Insect Biochemistry and Molecular Biology* **61**, 53-61.

Clarke, C. & Sheppard, P. M. 1975 Genetics of Mimetic Butterfly *Hypolimnas bolina* (L). *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **272**, 229-267.

Clarke, C., Sheppard, P. M. & Scali, V. 1975 All Female Broods in Butterfly *Hypolimnas bolina* (L). *Proceedings of the Royal Society B-Biological Sciences* **189**, 29-37.

Clarke, C. A. & Sheppard, P. M. 1959 The Genetics of *Papilio dardanus*, Brown. 1. Race Cenea from South Africa. *Genetics* **44**, 1347-1358.

Clarke, C. A. & Sheppard, P. M. 1964 Genetic Control of Melanic form insularia of Moth *Biston betularia* (L). *Nature* **202**, 215-216.

Clough, E., Jimenez, E., Kim, Y.-A., Whitworth, C., Neville, M. C., Hempel, L. U., Pavlou, H. J., Chen, Z.-X., Sturgill, D., Dale, R. K., Smith, H. E., Przytycka, T. M., Goodwin, S. F., Van Doren, M. & Oliver, B. 2014 Sex- and Tissue-Specific Functions of *Drosophila doublesex* Transcription Factor Target Genes. *Developmental Cell* **31**, 761-773.

Conte, G. L., Arnegard, M. E., Peichel, C. L. & Schluter, D. 2012 The Probability of Genetic Parallelism and Convergence in Natural Populations. *Proceedings of the Royal Society B-Biological Sciences* **279**, 5039-5047.

Cook, L. M. 2003 The Rise and Fall of the *Carbonaria* Form of the Peppered Moth. *Quarterly Review of Biology* **78**, 399-417.

Cook, L. M., Riley, A. M. & Woiwod, I. P. 2002 Melanic Frequencies in Three Species of Moths in Post Industrial Britain. *Biological Journal of the Linnean Society* **75**, 475-482.

Cordaux, R., Bouchon, D. & Greve, P. 2011 The Impact of Endosymbionts on the Evolution of Host Sex-Determination Mechanisms. *Trends in Genetics* **27**, 332-341.

Cordaux, R., Michel-Salzat, A., Frelon-Raimond, M., Rigaud, T. & Bouchon, D. 2004 Evidence for a New Feminizing *Wolbachia* Strain in the Isopod *Armadillidium vulgare*: Evolutionary Implications. *Heredity* **93**, 78-84.

Cosmides, L. M. & Tooby, J. 1981 Cytoplasmic Inheritance and Intragenomic Conflict. *Journal of Theoretical Biology* **89**, 83-129.

Cotter, W. B. 1974 Studies of Melanism in *Ephestia kühniella* Z. 1. Genetic and Physiological Characteristics of Mutant, *Ala nigra* (An). *Journal of Heredity* **65**, 213-216.

d'Alencon, E., Sezutsu, H., Legeai, F., Permal, E., Bernard-Samain, S., Gimenez, S., Gagneur, C., Cousserans, F., Shimomura, M., Brun-Barale, A., Flutre, T., Couloux, A., East, P., Gordon, K., Mita, K., Quesneville, H., Fournier, P. & Feyereisen, R. 2010 Extensive Synteny Conservation of Holocentric Chromosomes in Lepidoptera Despite High Rates of Local Genome Rearrangements. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 7680-7685.

Dai, F.-y., Qiao, L., Tong, X.-l., Cao, C., Chen, P., Chen, J., Lu, C. & Xiang, Z.-h. 2010 Mutations of an Arylalkylamine-N-acetyltransferase, Bm-iAANAT, are Responsible for Silkworm Melanism Mutant. *Journal of Biological Chemistry* **285**, 19553-19560.

Darwin, C. R. 1859 *The origin of species: by means of natural selection, or the preservation of favoured races in the struggle for life*. Cambridge: Cambridge University Press.

Darwin, C. R. 1871 The Descent of Man, and Selection in Relation to Sex.

Darwin, C. R. & Wallace, A. R. 1858 On the Tendency of Species to Form Varieties; and on the Perpetuation of Varieties and Species by Natural Means of Selection. *Journal of the Proceedings of the Linnean Society of London. Zoology* **3**, 45-50.

Dedeine, F., Vavre, F., Fleury, F., Loppin, B., Hochberg, M. E. & Bouletreau, M. 2001 Removing symbiotic *Wolbachia* Bacteria Specifically Inhibits Oogenesis in a Parasitic Wasp. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 6247-6252.

Devi, T. R., Amruthavalli, C. & Shyamala, B. V. 2013 Evolution of Sex Comb from the Primitive Bristle Pattern in *Drosophila* is Associated with Modification in the Developmental Regulatory Protein Dachshund. *Genesis* **51**, 97-109.

Duan, J., Li, R., Cheng, D., Fan, W., Zha, X., Cheng, T., Wu, Y., Wang, J., Mita, K., Xiang, Z. & Xia, Q. 2010 SilkDB v2.0: a Platform for Silkworm (*Bombyx mori*) Genome Biology. *Nucleic Acids Research* **38**, D453-D456.

Duan, J., Xu, H., Wang, F., Ma, S., Zha, X., Guo, H., Zhao, P. & Xia, Q. 2013a Novel Female-Specific Trans-Spliced and Alternative Splice Forms of *dsx* in the Silkworm *Bombyx mori*. *Biochemical and biophysical research communications* **431**, 630-5.

Duan, J., Xu, H., Wang, F., Ma, S., Zha, X., Guo, H., Zhao, P. & Xia, Q. 2013b Novel Female-Specific Trans-Spliced and Alternative Splice Forms of *dsx* in the Silkworm *Bombyx mori*. *Biochemical and Biophysical Research Communications* **431**, 630-635.

Dunn, K. L. 2007 A Report on Butterfly Absence on Easter Island (Chile) During March. *Victorian Entomologist* **37**, 15-16.

Duron, O., Bouchon, D., Boutin, S., Bellamy, L., Zhou, L., Engelstaedter, J. & Hurst, G. D. 2008 The Diversity of Reproductive Parasites Among Arthropods: *Wolbachia* do not Walk Alone. *Bmc Biology* **6**, 27.

Dyson, E. A. & Hurst, G. D. D. 2004 Persistence of an Extreme Sex-Ratio Bias in a Natural Population. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 6520-6523.

Dyson, E. A., Kamath, M. K. & Hurst, G. D. D. 2002 *Wolbachia* Infection Associated with All-Female Broods in *Hypolimnas bolina* (Lepidoptera : Nymphalidae): Evidence for Horizontal Transmission of a Butterfly Male Killer. *Heredity* **88**, 166-171.

Eirin-Lopez, J. M. & Sanchez, L. 2015 The Comparative Study of Five Sex-Determining Proteins Across Insects Unveils High Rates of Evolution at Basal Components of the Sex Determination Cascade. *Development Genes and Evolution* **225**, 23-30.

Excoffier, L., Hofer, T. & Foll, M. 2009 Detecting Loci Under Selection in a Hierarchically Structured Population. *Heredity* **103**, 285-298.

Falconer, D. S. 1952 The Problem of Environment and Selection. *American Naturalist* **86**, 293-298.

Faure, M. F., David, P., Bonhomme, F. & Bierne, N. 2008 Genetic Hitchhiking in a Subdivided Population of *Mytilus edulis*. *Bmc Evolutionary Biology* **8**, 164.

Fay, J. C. & Wu, C. I. 2000 Hitchhiking Under Positive Darwinian Selection. *Genetics* **155**, 1405-1413.

Ferguson, L., Lee, S. F., Chamberlain, N., Nadeau, N., Joron, M., Baxter, S., Wilkinson, P., Papanicolaou, A., Kumar, S., Kee, T.-J., Clark, R., Davidson, C., Glithero, R., Beasley, H., Vogel, H., Ffrench-Constant, R. & Jiggins, C. 2010 Characterization of a Hotspot for Mimicry: Assembly of a Butterfly Wing Transcriptome to Genomic Sequence at the HmYb/Sb Locus. *Molecular Ecology* **19**, 240-254.

Ferguson, L. C., Maroja, L. & Jiggins, C. D. 2011 Convergent, Modular Expression of ebony and tan in the Mimetic Wing Patterns of *Heliconius* Butterflies. *Development Genes and Evolution* **221**, 297-308.

Ferree, P. M., Avery, A., Azpurua, J., Wilkes, T. & Werren, J. H. 2008 A Bacterium Targets Maternally Inherited Centrosomes to Kill Males in *Nasonia*. *Current Biology* **18**, 1409-1414.

Fisher, R. A. 1930 *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press.

Fisher, R. A. 1937 The Wave of Advance of Advantageous Genes. *Annals of Eugenics* **7**, 355-369.

Fujii, T. & Shimada, T. 2007 Sex Determination in the Silkworm, *Bombyx mori*: A Female Determinant on the W Chromosome and the Sex-Determining Gene Cascade. *Seminars in Cell & Developmental Biology* **18**, 379-388.

Fujii, Y., Kageyama, D., Hoshizaki, S., Ishikawa, H. & Sasaki, T. 2001 Transfection of *Wolbachia* in Lepidoptera: the Feminizer of the Adzuki Bean Borer *Ostrinia scapulalis* Causes Male Killing in the Mediterranean Flour Moth *Ephestia kuehniella*. *Proceedings of the Royal Society B-Biological Sciences* **268**, 855-859.

Fukui, T., Kawamoto, M., Shoji, K., Kiuchi, T., Sugano, S., Shimada, T., Suzuki, Y. & Katsuma, S. 2015 The Endosymbiotic Bacterium *Wolbachia* Selectively Kills Male Hosts by Targeting the Masculinizing Gene. *Plos Pathogens* **11**, e1005048.

Futahashi, R. & Fujiwara, H. 2005 Molecular Mechanisms of Melanization in the Larval Body Marking of a Swallowtail Butterfly, *Papilio xuthus*. *Mechanisms of Development* **122**, S155-S155.

Futahashi, R. & Fujiwara, H. 2007 Regulation of 20-hydroxyecdysone on the Larval Pigmentation and the Expression of Melanin Synthesis Enzymes and yellow Gene of the Swallowtail Butterfly, *Papilio xuthus*. *Insect Biochemistry and Molecular Biology* **37**, 855-864.

Futahashi, R., Sato, J., Meng, Y., Okamoto, S., Daimon, T., Yamamoto, K., Suetsugu, Y., Narukawa, J., Takahashi, H., Banno, Y., Katsuma, S., Shimada, T., Mita, K. & Fujiwara, H. 2008 yellow and ebony are the Responsible Genes for the Larval Color Mutants of the Silkworm *Bombyx Mori*. *Genetics* **180**, 1995-2005.

Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A. & Carroll, S. B. 2005 Chance Caught on the Wing: *cis*-Regulatory Evolution and the Origin of Pigment Patterns in *Drosophila*. *Nature* **433**, 481-487.

Gorman, M. J., An, C. & Kanost, M. R. 2007a Characterization of tyrosine hydroxylase from *Manduca sexta*. *Insect Biochemistry and Molecular Biology* **37**, 1327-1337.

Gorman, M. J., Wang, Y., Jiang, H. & Kanost, M. R. 2007b *Manduca sexta* Hemolymph Proteinase 21 Activates Prophenoloxidase-Activating Proteinase 3 in an Insect Innate Immune Response Proteinase Cascade. *Journal of Biological Chemistry* **282**, 11742-11749.

Gotoh, H., Miyakawa, H., Ishikawa, A., Ishikawa, Y., Sugime, Y., Emlen, D. J., Lavine, L. C. & Miura, T. 2014 Developmental Link Between Sex and Nutrition; *doublesex* Regulates Sex-Specific Mandible Growth via Juvenile Hormone Signaling in Stag Beetles. *Plos Genetics* **10**, e1004098.

Gould, S. J. 1989 *Wonderful Life: The Burgess Shale and the Nature of History*.

Gratz, S. J., Cummings, A. M., Nguyen, J. N., Hamm, D. C., Donohue, L. K., Harrison, M. M., Wildonger, J. & O'Connor-Giles, K. M. 2013 Genome Engineering of *Drosophila* with the CRISPR RNA-Guided Cas9 Nuclease. *Genetics* **194**, 1029-1035.

Gunderson, A. R., Frame, A. M., Swaddle, J. P. & Forsyth, M. H. 2008 Resistance of Melanized Feathers to Bacterial Degradation: is it Really so Black and White? *Journal of Avian Biology* **39**.

Haine, E. R. 2008 Symbiont-Mediated Protection. *Proceedings of the Royal Society B-Biological Sciences* **275**, 353-361.

Hamilton, W. D. 1967 Extraordinary Sex Ratios. *Science* **156**, 477-488.

Hanson, M. R. & Bentolila, S. 2004 Interactions of Mitochondrial and Nuclear Genes that Affect Male Gametophyte Development. *Plant Cell* **16**, S154-S169.

Hilgenboecker, K., Hammerstein, P., Schlattmann, P., Telschow, A. & Werren, J. H. 2008 How Many Species are Infected with *Wolbachia*? - a Statistical Analysis of Current Data. *Fems Microbiology Letters* **281**, 215-220.

Hiruma, K. & Riddiford, L. M. 2009 The Molecular Mechanisms of Cuticular Melanization: The Ecdysone Cascade Leading to dopa decarboxylase Expression in *Manduca sexta*. *Insect Biochemistry and Molecular Biology* **39**, 245-253.

Hoekstra, H. E., Drumm, K. E. & Nachman, M. W. 2004 Ecological Genetics of Adaptive Color Polymorphism in Pocket Mice: Geographic Variation in Selected and Neutral Genes. *Evolution* **58**, 1329-1341.

Hornett, E. A., Charlat, S., Duplouy, A. M. R., Davies, N., Roderick, G. K., Wedell, N. & Hurst, G. D. D. 2006 Evolution of Male-Killer Suppression in a Natural Population. *Plos Biology* **4**, 1643-1648.

Hornett, E. A., Charlat, S., Wedell, N., Jiggins, C. D. & Hurst, G. D. D. 2009 Rapidly Shifting Sex Ratio across a Species Range. *Current Biology* **19**, 1628-1631.

Hornett, E. A., Duplouy, A. M. R., Davies, N., Roderick, G. K., Wedell, N., Hurst, G. D. D. & Charlat, S. 2008 You Can't Keep a Good Parasite Down: Evolution of a Male-Killer Suppressor Uncovers Cytoplasmic Incompatibility. *Evolution* **62**, 1258-1263.

Hornett, E. A., Engelstadter, J. & Hurst, G. D. D. 2010 Hidden Cytoplasmic Incompatibility Alters the Dynamics of Male-Killer/Host Interactions. *Journal of Evolutionary Biology* **23**, 479-487.

Hornett, E. A., Moran, B., Reynolds, L. A., Charlat, S., Tazzyman, S., Wedell, N., Jiggins, C. D. & Hurst, G. D. D. 2014 The Evolution of Sex Ratio Distorter Suppression Affects a 25 cM Genomic Region in the Butterfly *Hypolimnas bolina*. *Plos Genetics* **10**, e1004822.

Hurst, G. D. D. & Frost, C. L. 2015 Reproductive Parasitism: Maternally Inherited Symbionts in a Biparental World. *Cold Spring Harbor Perspectives in Biology* **7**, a017699.

Hurst, G. D. D., Jiggins, F. M., von der Schulenburg, J. H. G., Bertrand, D., West, S. A., Goriacheva, II, Zakharov, I. A., Werren, J. H., Stouthamer, R. & Majerus, M. E. N. 1999

Male-killing *Wolbachia* in Two Species of Insect. *Proceedings of the Royal Society B-Biological Sciences* **266**, 735-740.

Ito, K., Katsuma, S., Kuwazaki, S., Jouraku, A., Fujimoto, T., Sahara, K., Yasukochi, Y., Yamamoto, K., Tabunoki, H., Yokoyama, T., Kadono-Okuda, K. & Shimada, T. 2016 Mapping and Recombination Analysis of Two Moth Colour Mutations, Black Moth and Wild Wing Spot, in the Silkworm *Bombyx mori*. *Heredity* **116**, 52-59.

Ito, Y., Harigai, A., Nakata, M., Hosoya, T., Araya, K., Oba, Y., Ito, A., Ohde, T., Yaginuma, T. & Niimi, T. 2013 The Role of *doublesex* in the Evolution of Exaggerated Horns in the Japanese Rhinoceros Beetle. *Embo Reports* **14**, 561-567.

Jaenike, J. 1996 Sex-Ratio Meiotic Drive in the *Drosophila quinaria* Group. *American Naturalist* **148**, 237-254.

Jaenike, J. 1999 Suppression of Sex-Ratio Meiotic Drive and the Maintenance of Y-Chromosome Polymorphism in *Drosophila* (vol 53, pg 164, 1999). *Evolution* **53**, 1326-1326.

Jaenike, J. 2007a Fighting Back Against Male-Killers. *Trends in Ecology & Evolution* **22**, 167-169.

Jaenike, J. 2007b Spontaneous Emergence of a New *Wolbachia* Phenotype. *Evolution* **61**, 2244-2252.

Jiggins, F. M., Bentley, J. K., Majerus, M. E. N. & Hurst, G. D. D. 2001 How Many Species are Infected with *Wolbachia*? Cryptic Sex Ratio Distorters Revealed to be Common by Intensive Sampling. *Proceedings of the Royal Society B-Biological Sciences* **268**, 1123-1126.

Jiggins, F. M., Hurst, G. D. D. & Majerus, M. E. N. 1999 How Common are Meiotically Driving Sex Chromosomes in Insects? *American Naturalist* **154**, 481-483.

Joron, M., Frezal, L., Jones, R. T., Chamberlain, N. L., Lee, S. F., Haag, C. R., Whibley, A., Becuwe, M., Baxter, S. W., Ferguson, L., Wilkinson, P. A., Salazar, C., Davidson, C., Clark, R., Quail, M. A., Beasley, H., Glithero, R., Lloyd, C., Sims, S., Jones, M. C., Rogers, J., Jiggins, C. D. & Ffrench-Constant, R. H. 2011 Chromosomal Rearrangements Maintain a Polymorphic Supergene Controlling Butterfly Mimicry. *Nature* **477**, 203-U102.

Joron, M., Papa, R., Beltran, M., Chamberlain, N., Mavarez, J., Baxter, S., Abanto, M., Bermingham, E., Humphray, S. J., Rogers, J., Beasley, H., Barlow, K., ffrench-Constant, R. H., Mallet, J., McMillan, W. O. & Jiggins, C. D. 2006 A Conserved Supergene Locus Controls Colour Pattern Diversity in *Heliconius* Butterflies. *Plos Biology* **4**, 1831-1840.

Kaplan, N. L., Hudson, R. R. & Langley, C. H. 1989 The Hitchhiking Effect Revisited. *Genetics* **123**, 887-899.

Kawamoto, M., Koga, H., Kiuchi, T., Shoji, K., Sugano, S., Shimada, T., Suzuki, Y. & Katsuma, S. 2015 Sexually Biased Transcripts at Early Embryonic Stages of the Silkworm Depend on the Sex Chromosome Constitution. *Gene* **560**, 50-56.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. & Drummond, A. 2012 Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data. *Bioinformatics* **28**, 1647-1649.

Kern, P., Cook, J. M., Kageyama, D. & Riegler, M. 2015 Double Trouble: Combined Action of Meiotic Drive and *Wolbachia* Feminization in *Eurema* Butterflies. *Biology Letters* **11**, 20150095.

Kettlewell, H. B. D. 1973 *The Evolution of Melanism*. Oxford: Clarendon Press.

Kijimoto, T., Moczek, A. P. & Andrews, J. 2012 Diversification of *doublesex* Uunction underlies Morph-, Sex-, and Species-Specific Development of Beetle Horns. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 20526-20531.

Kim, Y. 2013 Stochastic patterns of polymorphism after a selective sweep over a subdivided population. *Genetics Research* **95**, 57-67.

Kim, Y. & Maruki, T. 2011 Hitchhiking Effect of a Beneficial Mutation Spreading in a Subdivided Population. *Genetics* **189**, 213-226.

Kim, Y. & Nielsen, R. 2004 Linkage Disequilibrium as a Signature of Selective Sweeps. *Genetics* **167**, 1513-1524.

Kim, Y. & Stephan, W. 2002 Detecting a Local Signature of Genetic Hitchhiking Along a Recombining Chromosome. *Genetics* **160**, 765-777.

Kiuchi, T., Koga, H., Kawamoto, M., Shoji, K., Sakai, H., Arai, Y., Ishihara, G., Kawaoka, S., Sugano, S., Shimada, T., Suzuki, Y., Suzuki, M. G. & Katsuma, S. 2014 A Single Female-Specific piRNA is the Primary Determiner of Sex in the Silkworm. *Nature* **509**, 633-636.

Klopfstein, S., Currat, M. & Excoffier, L. 2006 The Fate of Mutations Surfing on the Wave of a Range Expansion. *Molecular Biology and Evolution* **23**, 482-490.

Koch, P. B., Behnecke, B. & ffrench-Constant, R. H. 2000a The Molecular Basis of Melanism and Mimicry in a Swallowtail Butterfly. *Current Biology* **10**, 591-594.

Koch, P. B., Behnecke, B., Weigmann-Lenz, M. & Ffrench-Constant, R. H. 2000b Insect Pigmentation: Activities of beta-alanyldopamine Synthase in Wing Color Patterns of Wild-Type and Melanic Mutant Swallowtail Butterfly *Papilio glaucus*. *Pigment Cell Research* **13**, 54-58.

Koch, P. B., Keys, D. N., Rocheleau, T., Aronstein, K., Blackburn, M., Carroll, S. B. & Ffrench-Constant, R. H. 1998 Regulation of dopa decarboxylase Expression During Colour Pattern Formation in Wild-Type and Melanic Tiger Swallowtail Butterflies. *Development* **125**, 2303-2313.

Koehn, R. K., Newell, R. I. E. & Immermann, F. 1980 Maintenance of an Aminopeptidase Allele Frequency Cline by Natural Selection. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **77**, 5385-5389.

Kronforst, M. R., Kapan, D. D. & Gilbert, L. E. 2006 Parallel Genetic Architecture of Parallel Adaptive Radiations in Mimetic *Heliconius* Butterflies. *Genetics* **174**, 535-539.

Kuhn, A. & Henke. 1929 Genetische und Entwicklungsphysiologische Untersuchungen an der Mehlmotte *Ephestia Kuehniella* Zeller. I-VII *Abhandlungen der Gesellschaft der Wissenschaften in Göttingen, Mathematisch-Physikalische* **15**, 1-151.

Kumar, S., Stecher, G. & Tamura, K. 2016 MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution* **33**, 1870-1874.

Kunte, K., Zhang, W., Tenger-Trolander, A., Palmer, D. H., Martin, A., Reed, R. D., Mullen, S. P. & Kronforst, M. R. 2014 *doublesex* is a Mimicry Supergene. *Nature* **507**, 229-232.

Lee, Y. W., Gould, B. A. & Stinchcombe, J. R. 2014 Identifying the Genes Underlying Quantitative Traits: a Rationale for the QTN Programme. *Aob Plants* **6**, plu004.

Lees, D., R. 1971 The Distribution of Melanism in the Pale Bridled Beauty Moth, *Phigalia pedaria*, in Great Britain. In *Ecological Genetics and Evolution* (ed. E. R. Creed), pp. 152-174: Blackwell, Oxford.

Lees, D. R. 1974 Genetic Control of Melanic Forms of Moth *Phigalia pilosaria* (*pedaria*). *Heredity* **33**, 145-150.

Lees, D. R., Creed, E. R. & Duckett, J. G. 1973 Atmospheric Pollution and Industrial Melanism. *Heredity* **30**, 227-232.

Lenormand, T., Guillemaud, T., Bourguet, D. & Raymond, M. 1998 Appearance and Sweep of a Gene Duplication: Adaptive Response and Potential for New Functions in the Mosquito *Culex pipiens*. *Evolution* **52**, 1705-1712.

Lindholm, A. K., Dyer, K. A., Firman, R. C., Fishman, L., Forstmeier, W., Holman, L., Johannesson, H., Knief, U., Kokko, H., Larracuente, A. M., Manser, A., Montchamp-Moreau, C., Petrosyan, V. G., Pomiankowski, A., Presgraves, D. C., Safronova, L. D., Sutter, A., Unckless, R. L., Verspoor, R. L., Wedell, N., Wilkinson, G. S. & Price, T. A. R. 2016 The Ecology and Evolutionary Dynamics of Meiotic Drive. *Trends in Ecology & Evolution* **31**, 315-326.

Liu, S., Wang, M. & Li, X. 2015 Overexpression of Tyrosine hydroxylase and Dopa decarboxylase Associated with Pupal Melanization in *Spodoptera exigua*. *Scientific Reports* **5**, 11273.

Lu, C., Dai, F. & Xiang, Z. 2003 Studies on the Mutation Strains of the *Bombyx mori* Gene Bank. *Scientia Agricultura Sinica* **36**, 968-975.

Luo, S. D. & Baker, B. S. 2015 Constraints on the Evolution of a *doublesex* Target Gene Arising from *doublesex's* Pleiotropic Deployment. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E852-E861.

Luo, S. D., Shi, G. W. & Baker, B. S. 2011 Direct Targets of the *D. melanogaster* DSXF Protein and the Evolution of Sexual Development. *Development* **138**, 2761-2771.

Maeda, T. 1939 Chiasma Studies in the Silkworm *Bombyx mori* L. *Japanese Journal of Genetics* **15**, 118-127.

Majerus, M. E. N. 1998 Melanism: Evolution in action. *Melanism: Evolution in action*.

Majerus, T. M. O. & Majerus, M. E. N. 2010a Intergenomic Arms Races: Detection of a Nuclear Rescue Gene of Male-Killing in a Ladybird. *Plos Pathogens* **6**, e1000987.

Majerus, T. M. O. & Majerus, M. E. N. 2010b Intergenomic Arms Races: Detection of a Nuclear Rescue Gene of Male-Killing in a Ladybird. *PLoS pathogens* **6**, e1000987-e1000987.

Marec, F. 1990 Genetic Control of Pest Lepidoptera - Induction of Sex-Linked Recessive Lethal Mutations in *Ephestia kuehniella* (Pyralidae). *Acta Entomologica Bohemoslovaca* **87**, 445-458.

Marec, F. 1991 Genetic Control of Pest Lepidoptera Construction of a Balanced Lethal Strain in *Ephestia kuehniella*. *Entomologia Experimentalis Et Applicata* **61**, 271-283.

Marec, F., Tothova, A., Sahara, K. & Traut, W. 2001 Meiotic Pairing of Sex Chromosome, Fragments and its Relation to Atypical Transmission of a Sex-Linked Marker in *Ephestia kuehniella* (Insecta : Lepidoptera). *Heredity* **87**, 659-671.

Marec, F. & Traut, W. 1994 Sex-Chromosome Pairing and Sex-Chromatin bodies in W-Z Translocation Strains of *Ephestia kuehniella* (Lepidoptera). *Genome* **37**, 426-435.

Martin, A. & Orgogozo, V. 2013 The Loci of Repeated Evolution: A Catalog of Genetic Hotspots of Phenotypic Variation. *Evolution* **67**, 1235-1250.

Mather, K. 1950 The Genetical Architecure of Heterostyly in *Primula sinensis*. *Evolution* **4**, 340-352.

McVean, G. 2007 The Structure of Linkage Disequilibrium Around a Selective Sweep. *Genetics* **175**, 1395-1406.

Mercot, H., Atlan, A., Jacques, M. & Montchampmoreau, C. 1995 Sex Ratio Distortion in *Drosophila simulans* - Cooccurrence of a Meiotic Drive and a Suppressor of Drive. *Journal of Evolutionary Biology* **8**, 283-300.

Merrill, R. M., Van Schooten, B., Scott, J. A. & Jiggins, C. D. 2011 Pervasive Genetic Associations Between Traits Causing Reproductive Isolation in *Heliconius* Butterflies. *Proceedings of the Royal Society B-Biological Sciences* **278**, 511-518.

Mitsuhashi, W., Fukuda, H., Nicho, K. & Murakami, R. 2004 Male-killing *Wolbachia* in the Butterfly *Hypolimnas bolina*. *Entomologia Experimentalis Et Applicata* **112**, 57-64.

Mitsuhashi, W., Ikeda, H. & Muraji, M. 2011 Fifty-Year Trend Towards Suppression of *Wolbachia*-Induced Male-Killing by its Butterfly Host, *Hypolimnas bolina. Journal of Insect Science* **11**, 92.

Moran, N. A., McCutcheon, J. P. & Nakabachi, A. 2008 Genomics and Evolution of Heritable Bacterial Symbionts. In *Annual Review of Genetics*, vol. 42, pp. 165-190. Palo Alto: Annual Reviews.

Morris, S. C. 2010 Evolution: Like any Other Science it is Predictable. *Philosophical Transactions of the Royal Society B-Biological Sciences* **365**, 133-145.

Mueller, C., Mytnik, L. & Quastel, J. 2011 Effect of Noise on Front Propagation in Reaction-Diffusion Equations of KPP Type. *Inventiones Mathematicae* **184**, 405-453.

Nappi, A. J. & Christensen, B. M. 2005 Melanogenesis and Associated Cytotoxic Reactions: Applications to Insect Innate Immunity. *Insect Biochemistry and Molecular Biology* **35**, 443-459.

Nielsen, R. 2005 Molecular Signatures of Natural Selection. *Annual Review of Genetics* **39**, 197-218.

Nishikawa, H., Iijima, T., Kajitani, R., Yamaguchi, J., Ando, T., Suzuki, Y., Sugano, S., Fujiyama, A., Kosugi, S., Hirakawa, H., Tabata, S., Ozaki, K., Morimoto, H., Ihara, K., Obara, M., Hori, H., Itoh, T. & Fujiwara, H. 2015 A Genetic Mechanism for Female-Limited Batesian Mimicry in *Papilio* Butterfly. *Nature Genetics* **47**, 405-U169.

Ohbayashi, F., Suzuki, M. G., Mita, K., Okano, K. & Shimada, T. 2001 A Homologue of the *Drosophila doublesex* Gene is Transcribed into Sex-Specific mRNA Isoforms in the Silkworm, *Bombyx mori*. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology* **128**, 145-158.

Ohbayashi, F., Suzuki, M. G. & Shimada, T. 2002 Sex Determination in *Bombyx mori*. *Current Science* **83**, 466-471.

Papa, R., Martin, A. & Reed, R. D. 2008 Genomic Hotspots of Adaptation in Butterfly Wing Pattern Evolution. *Current Opinion in Genetics & Development* **18**, 559-564.

Pardo-Diaz, C., Salazar, C. & Jiggins, C. D. 2015 Towards the Identification of the Loci of Adaptive Evolution. *Methods in Ecology and Evolution* **6**, 445-464.

Patrick, B. H. 2004 Invasion of Blue Moon Butterflies in Taranaki. *Weta* **28**, 45-46.

Poulton, E. B. 1923 All-Female Families of *Hypolimnas bolina* L., bred in Fiji by H.W. Simmonds. *Transactions of the Royal Entomological Society of London* **71**, ix-xii.

Poulton, E. B. 1924 Mimicry in the Butterflies of Fiji Considered in Relation to the *Euplorine* and *Danaineinvasions* of Polynesia and to the Female Forms of *Hypolimnas bolina* L., in the Pacific. *Transactions of the Royal Entomological Society of London* **71**, 566-691.

Poulton, E. B. a. R., N.D. 1928 The Rhopalocera of the "St George" Expedition from French Oceania. *Transactions of the Royal Entomological Society of London* **76**, 453-463.

Price, D. C., Egizi, A. & Fonseca, D. M. 2015 The Ubiquity and Ancestry of Insect *doublesex*. *Scientific Reports* **5**, 13068.

Pringle, E. G., Baxter, S. W., Webster, C. L., Papanicolaou, A., Lee, S. F. & Jiggins, C. D. 2007 Synteny and Chromosome Evolution in the Lepidoptera: Evidence from Mapping in *Heliconius melpomene*. *Genetics* **177**, 417-426.

Prud'homme, B., Gompel, N., Rokas, A., Kassner, V. A., Williams, T. M., Yeh, S. D., True, J. R. & Carroll, S. B. 2006 Repeated Morphological Evolution Through *cis*-Regulatory Changes in a Pleiotropic Gene. *Nature* **440**, 1050-1053.

Przeworski, M. 2002 The Signature of Positive Selection at Randomly Chosen Loci. *Genetics* **160**, 1179-1189.

Rahman, M. M., Roberts, H. L. S., Sarjan, M., Asgari, S. & Schmidt, O. 2004 Induction and Transmission of *Bacillus thuringiensis* Tolerance in the Flour Moth *Ephestia kuehniella*. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 2696-2699.

Ramsay, G. W. 1971 The Blue Moon Butterfly *Hypolimnas bolina nerina* in New Zealand During Autumn 1971. *New Zealand Entomologist* **5**, 73-75.

Ramsay, G. W. & Ordish, R. G. 1966 Australian Blue Moon Butterfly *Hypolimnas bolina nerina* (F) in New Zealand. *New Zealand Journal of Science* **9**, 719-&.

Randerson, J. P., Smith, N. G. C. & Hurst, L. D. 2000 The Evolutionary Dynamics of Male-Killers and Their Hosts. *Heredity* **84**, 152-160.

Rausher, M. D. & Delph, L. F. 2015 Commentary: When does Understanding Phenotypic Evolution Require Identification of the Underlying Genes? *Evolution* **69**, 1655-1664.

Raymond, M. & Rousset, F. 1995 Genepop (Version 1.2) - Population Genetics Software for Exact Tests and Ecumenicism. *Journal of Heredity* **86**, 248-249.

Regnart, H. C. 1933 Additions to our Knowledge of Chromosome Numbers in the Lepidoptera. *Proceedings of the University of Durham Philosophical Society* **9**, 79-83.

Rigaud, T. & Juchault, P. 1992 Genetic Control of the Vertical Transmission of a Cytoplasmic Sex Factor in *Armadillidium vulgare* latr (Crustacea, Oniscidea). *Heredity* **68**, 47-52.

Rigaud, T. & Juchault, P. 1993 Conflict Between Feminizing Sex-Ratio Distorters and an Autosomal Masculinizing Gene in the Terrestrial Isopod *Armadillidium vulgare* Latr. *Genetics* **133**, 247-252.

Rigaud, T., Juchault, P. & Mocquard, J. P. 1997 The Evolution of Sex Determination in Isopod Crustaceans. *Bioessays* **19**, 409-416.

Rigaud, T., Moreau, J. & Juchault, P. 1999 *Wolbachia* Infection in the Terrestrial Isopod *Oniscus asellus*: Sex Ratio Distortion and Effect on Fecundity. *Heredity* **83**, 469-475.

Riparbelli, M. G., Giordano, R., Ueyama, M. & Callaini, G. 2012 *Wolbachia*-Mediated Male Killing Is Associated with Defective Chromatin Remodeling. *Plos One* **7**, e30045.

Robinson, R. 1971 *Lepidoptera genetics*. Oxford: Pergamon Press.

Rockman, M. V. 2012 The QTN Program and the Alleles that Matter for Evolution: All that's Gold does Not Glitter. *Evolution* **66**, 1-17.

Rosenberg, D. R. 1973 The 1st Breeding Record of the Blue Moon Butterfly Hypolimnas bolina nerina in New Zealand. *New Zealand Entomologist* **5**, 191-192.

Rousset, F. 2008 GENEPOP ' 007: a Complete Re-implementation of the GENEPOP Software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z. P., Richter, D. J., Schaffner, S. F., Gabriel, S. B., Platko, J. V., Patterson, N. J., McDonald, G. J., Ackerman, H. C., Campbell, S. J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R. & Lander, E. S. 2002 Detecting Recent Positive Selection in the Human Genome from Haplotype Structure. *Nature* **419**, 832-837.

Saccheri, I. J., Rousset, F., Watts, P. C., Brakefield, P. M. & Cook, L. M. 2008 Selection and Gene Flow on a Diminishing Cline of Melanic Peppered Moths. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 16212-16217.

Saenko, S. V., Brakefield, P. M. & Beldade, P. 2010 Single Locus Affects Embryonic Segment Polarity and Multiple Aspects of an Adult Evolutionary Novelty. *Bmc Biology* **8**, 111.

Sahara, K., Marec, F. & Traut, W. 1999 TTAGG Telomeric Repeats in Chromosomes of Some Insects and other Arthropods. *Chromosome Research* **7**, 449-460.

Sakamoto, H., Ishikawa, Y., Sasaki, T., Kikuyama, S., Tatsuki, S. & Hoshizaki, S. 2005 Transinfection Reveals the Crucial Importance of *Wolbachia* Genotypes in Determining the Type of Reproductive Alteration in the Host. *Genetical Research* **85**, 205-210.

Santiago, E. & Caballero, A. 2005 Variation After a Selective Sweep in a Subdivided population. *Genetics* **169**, 475-483.

Sasaki, T., Kubo, T. & Ishikawa, H. 2002 Interspecific Transfer of *Wolbachia* Between Two Lepidopteran Insects Expressing Cytoplasmic Incompatibility: A *Wolbachia* Variant

Naturally Infecting *Cadra cautella* Causes Male Killing in *Ephestia kuehniella*. *Genetics* **162**, 1313-1319.

Scriber, J. M., Hagen, R. H. & Lederhouse, R. C. 1996 Genetics of Mimicry in the Tiger Swallowtail Butterflies, *Papilio glaucus* and P-canadensis (Lepidoptera: Papilionidae). *Evolution* **50**, 222-236.

Shirangi, T. R., Dufour, H. D., Williams, T. M. & Carroll, S. B. 2009 Rapid Evolution of Sex Pheromone-Producing Enzyme Expression in *Drosophila*. *Plos Biology* **7,** e1000168.

Shukla, J. N. & Nagaraju, J. 2010 Two Female-Specific DSX Proteins are Encoded by the Sex-Specific Transcripts of *dsx*, and are Required for Female Sexual Differentiation in Two Wild Silkmoth Species, *Antheraea assama* and *Antheraea mylitta* (Lepidoptera, Saturniidae). *Insect Biochemistry and Molecular Biology* **40**, 672-682.

Simmonds, H. W. 1926 Sex-ratio of *Hypolimnas bolina*, L. in Viti Leuvu, Fiji. *Proceedings of the Entomological Society of London* **1**, 29-31.

Simmonds, H. W. 1931 Further Notes on *Hyplimnas bolina* L., in Fiji. *The Proceedings of The Entomological Society of London* **5**, 75-77.

Siwicki, K. K. & Kravitz, E. A. 2009 Fruitless, *doublesex* and the Genetics of Social Behavior in *Drosophila melanogaster*. *Current Opinion in Neurobiology* **19**, 200-206.

Slatkin, M. & Wiehe, T. 1998 Genetic Hitch-Hiking in a Subdivided Population. *Genetics Research* **71**, 155-160.

Smith, J. M. & Haigh, J. 1974 The Hitchhiking Effect of a Favorable Gene. *Genetical Research* **23**, 23-35.

Sobrinho, I. S., Jr. & de Brito, R. A. 2012 Positive and Purifying Selection Influence the Evolution of *doublesex* in the *Anastrepha fraterculus* Species Group. *Plos One* **7,** e33446.

Stam, P. 1993 Construction of Integrated Genetic Linkage Maps by Means of a New Computer Package - Joinmap. *Plant Journal* **3**, 739-744.

Stephan, W., Song, Y. S. & Langley, C. H. 2006 The Hitchhiking Effect on Linkage Disequilibrium Between Linked Neutral Loci. *Genetics* **172**, 2647-2663.

Stephens, M. & Scheet, P. 2005 Accounting for Decay of Linkage Disequilibrium in Haplotype Inference and Missing-Data Imputation. *American Journal of Human Genetics* **76**, 449-462.

Stephens, M., Smith, N. J. & Donnelly, P. 2001 A New Statistical Method for Haplotype Reconstruction from Population Data. *American Journal of Human Genetics* **68**, 978-989.

Stern, D. L. & Orgogozo, V. 2008 The Loci of Evolution: How Predictable is Genetic Evolution? *Evolution* **62**, 2155-2177.

Stern, D. L. & Orgogozo, V. 2009 Is Genetic Evolution Predictable? *Science* **323**, 746-751.

Sugimoto, T. N., Fujii, T., Kayukawa, T., Sakamoto, H. & Ishikawa, Y. 2010 Expression of a *doublesex* Homologue is Altered in Sexual Mosaics of *Ostrinia scapulalis* Moths Infected with *Wolbachia*. *Insect Biochemistry and Molecular Biology* **40**, 847-854.

Sugimoto, T. N. & Ishikawa, Y. 2012 A Male-Killing *Wolbachia* Carries a Feminizing Factor and is Associated with Degradation of the Sex-Determining System of its Host. *Biology Letters* **8**, 412-415.

Sugimoto, T. N., Kayukawa, T., Shinoda, T., Ishikawa, Y. & Tsuchida, T. 2015 Misdirection of Dosage Compensation Underlies Bidirectional Sex-Specific Death in *Wolbachia*-Infected *Ostrinia scapulalis*. *Insect Biochemistry and Molecular Biology* **66**, 72-76.

Sugumaran, M. 2009 Complexities of Cuticular Pigmentation in Insects. *Pigment Cell & Melanoma Research* **22**, 523-525.

Suomalai.E, Cook, L. M. & Turner, J. R. G. 1973 Achiasmatic Oogenesis in Heliconiine Butterflies. *Hereditas* **74**, 302-304.

Tajima, F. 1989 Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics* **123**, 585-595.

Talloen, W., Van Dyck, H. & Lens, L. 2004 The Cost of Melanization: Butterfly Wing Coloration Under Environmental Stress. *Evolution* **58**, 360-366.

Tanaka, K., Barmina, O., Sanders, L. E., Arbeitman, M. N. & Kopp, A. 2011 Evolution of Sex-Specific Traits through Changes in HOX-Dependent *doublesex* Expression. *Plos Biology* **9**, e1001131.

Tao, Y., Araripe, L., Kingan, S. B., Ke, Y., Xiao, H. & Hartl, D. L. 2007a A Sex-Ratio Meiotic Drive System in *Drosophila simulan*s. II: An X-Linked Distorter. *Plos Biology* **5**, 2576-2588.

Tao, Y., Hartl, D. L. & Laurie, C. C. 2001 Sex-Ratio Segregation Distortion Associated with Reproductive Isolation in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 13183-13188.

Tao, Y., Masly, J. P., Araripe, L., Ke, Y. & Hartl, D. L. 2007b A Sex-Ratio Meiotic Drive System in *Drosophila simulans*. I: An Autosomal Suppressor. *Plos Biology* **5**, 2560-2575.

Tennent, W. J. 2006 A Checklist of the Butterflies of Melanesia, Micronesia, Polynesia and Some Adjacent Areas. *Zootaxa* **1178**.

Teotonio, H., Chelo, I. M., Bradic, M., Rose, M. R. & Long, A. D. 2009 Experimental Evolution Reveals Natural Selection on Standing Genetic Variation. *Nature Genetics* **41**, 251-257.

Traut, W. 1977 Study of Recombination, Formation of Chiasmata and Synaptonemal Complexes in Female and Male Meiosis of *Ephestia kuehniella* (Lepidoptera). *Genetica* **47**, 135-142.

Traut, W. & Marec, F. 1996 Sex Chromatin in Lepidoptera. *Quarterly Review of Biology* **71**, 239-256.

Traut, W. & Marec, F. 1997 Sex Chromosome Differentiation in Some Species of Lepidoptera (Insecta). *Chromosome Research* **5**, 283-291.

Traut, W., Sahara, K. & Marec, F. 2007 Sex Chromosomes and Sex Determination in Lepidoptera. *Sexual Development* **1**, 332-346.

Travisano, M. & Shaw, R. G. 2013 Lost in the Map. *Evolution* **67**, 305-314.

True, J. R. 2003 Insect Melanism: the Molecules Matter. *Trends in Ecology & Evolution* **18**, 640-647.

Truman, J. W., Riddifor.Lm & Safranek, L. 1973 Hormonal Control of Cuticle Coloration in Tobacco Hornworm, *Manduca sexta* - Basis of an Ultrasensitive Bioassay for Juvenile Hormone. *Journal of Insect Physiology* **19**, 195-203.

Turner, J. R. G. & Sheppard, P. M. 1975 Absence of Crossing-Over in Female Butterflies (Heliconius). *Heredity* **34**, 265-269.

van't Hof, A. E., Edmonds, N., Dalikova, M., Marec, F. & Saccheri, I. J. 2011 Industrial Melanism in British Peppered Moths Has a Singular and Recent Mutational Origin. *Science* **332**, 958-960.

Van't Hof, A. E., Nguyen, P., Dalikova, M., Edmonds, N., Marec, F. & Saccheri, I. J. 2013 Linkage map of the peppered moth, *Biston betularia* (Lepidoptera, Geometridae): a Model of Industrial Melanism. *Heredity* **110**, 283-295.

van't Hof, A. E. & Saccheri, I. J. 2010 Industrial Melanism in the Peppered Moth Is Not Associated with Genetic Variation in Canonical Melanisation Gene Candidates. *Plos One* **5**, e10889.

Veneti, Z., Bentley, J. K., Koana, T., Braig, H. R. & Hurst, G. D. D. 2005 A Functional Dosage Compensation Complex Required for Male Killing in *Drosophila*. *Science* **307**, 1461-1463.

Ventura, I. M., Martins, A. B., Lyra, M. L., Andrade, C. A. C., Carvalho, K. A. & Klaczko, L. B. 2012 *Spiroplasma* in *Drosophila melanogaster* Populations: Prevalence, Male-Killing, Molecular Identification, and No Association with *Wolbachia*. *Microbial Ecology* **64**, 794-801.

Verhulst, E. C. & van de Zande, L. 2015 Double nexus-*doublesex* is the Connecting Element in Sex Determination. *Briefings in Functional Genomics* **14**, 396-406.

Voight, B. F., Kudaravalli, S., Wen, X. Q. & Pritchard, J. K. 2006 A Map of Recent Positive Selection in the Human Genome. *Plos Biology* **4**, 446-458.

Voorrips, R. E. 2002 MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs. *Journal of Heredity* **93**, 77-78.

Walsberg, G. E. 1983 Avian Ecological Energetics. In *Avian Biology*, pp. 161-220. New York: Academic Press.

Wang, J., Xia, Q. Y., He, X. M., Dai, M. T., Ruan, J., Chen, J., Yu, G., Yuan, H. F., Hu, Y. F., Li, R. Q., Feng, T., Ye, C., Lu, C., Li, S. G., Wong, G. K. S., Yang, H. M., Xiang, Z. H., Zhou, Z. Y. & Yu, J. 2005 SilkDB: a Knowledgebase for Silkworm Biology and Genomics. *Nucleic Acids Research* **33**, D399-D402.

Wang, X.-Y., Zheng, Z.-Z., Song, H.-S. & Xu, Y.-Z. 2014 Conserved RNA cis-elements regulate Alternative Splicing of Lepidopteran *doublesex*. *Insect Biochemistry and Molecular Biology* **44**, 1-11.

Wang, Z. H., Zou, Y. J., Li, X. Y., Zhang, Q. Y., Chen, L., Wu, H., Su, D. H., Chen, Y. L., Guo, J. X., Luo, D., Long, Y. M., Zhong, Y. & Liu, Y. G. 2006 Cytoplasmic Male Sterility of Rice with Boro II Cytoplasm is Caused by a Cytotoxic Peptide and is Restored by Two Related PPR Motif Genes via Distinct Modes of mRNA Silencing. *Plant Cell* **18**, 676-687.

White, F. B. 1876 On Melanochroism and Leucochroism. *The Entomologist's Monthly Magazine* **13**, 145-149.

Williams, T. M., Selegue, J. E., Werner, T., Gompel, N., Kopp, A. & Carroll, S. B. 2008 The Regulation and Evolution of a Genetic Switch Controlling Sexually Dimorphic Traits in *Drosophila*. *Cell* **134**, 610-623.

Winstanley, W. J. 1982 Some Insects in the Rangitoto Range, New Zealand. *New Zealand Entomologist* **7**, 323-324.

Wittkopp, P. J. & Beldade, P. 2009 Development and Evolution of Insect Pigmentation: Genetic Mechanisms and the Potential Consequences of Pleiotropy. *Seminars in Cell & Developmental Biology* **20**, 65-71.

Wittkopp, P. J., Carroll, S. B. & Kopp, A. 2003a Evolution in Black and White: Genetic Control of Pigment Patterns in *Drosophila*. *Trends in Genetics* **19**, 495-504.

Wittkopp, P. J., True, J. R. & Carroll, S. B. 2002a Reciprocal Functions of the *Drosophila* Yellow and Ebony Proteins in the Development and Evolution of Pigment Patterns. *Development* **129**, 1849-1858.

Wittkopp, P. J., Vaccaro, K. & Carroll, S. B. 2002b Evolution of yellow Gene Regulation and Pigmentation in *Drosophila*. *Current Biology* **12**, 1547-1556.

Wittkopp, P. J., Williams, B. L., Selegue, J. E. & Carroll, S. B. 2003b *Drosophila* Pigmentation Evolution: Divergent Genotypes Underlying Convergent Phenotypes. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 1808-1813.

Wright, S. 1931 Evolution in Mendelian populations. *Genetics* **16**, 97-159.

Wright, S. 1932 The Roles of Mutation, Inbreeding, Crossbreeding and Selection in Evolution. *Proceedings of the 6th International Congress in Genetics* **1**, 356-366.

Wright, T. R. F. 1987 The Genetics of Biogenic Amine Metabolism, Sclerotization and Melanization in *Drosophila melanogaster*. *Advances in Genetics Incorporating Molecular Genetic Medicine* **24**, 127-222.

Xia, Q. Y., Zhou, Z. Y., Lu, C., Cheng, D. J., Dai, F. Y., Li, B., Zhao, P., Zha, X. F., Cheng, T. C., Chai, C. L., Pan, G. Q., Xu, J. S., Liu, C., Lin, Y., Qian, J. F., Hou, Y., Wu, Z. L., Li, G. R., Pan, M. H., Li, C. F., Shen, Y. H., Lan, X. Q., Yuan, L. W., Li, T., Xu, H. F., Yang, G. W., Wan, Y. J., Zhu, Y., Yu, M. D., Shen, W. D., Wu, D. Y., Xiang, Z. H., Yu, J., Wang, J., Li, R. Q., Shi, J. P., Li, H., Li, G. Y., Su, J. N., Wang, X. L., Li, G. Q., Zhang, Z. J., Wu, Q. F., Li, J., Zhang, Q. P., Wei, N., Xu, J. Z., Sun, H. B., Dong, L., Liu, D. Y., Zhao, S. L., Zhao, X. L., Meng, Q. S., Lan, F. D., Huang, X. G., Li, Y. Z., Fang, L., Li, D. W., Sun, Y. Q., Zhang, Z. P., Yang, Z., Huang, Y. Q., Xi, Y., Qi, Q. H., He, D. D., Huang, H. Y., Zhang, X. W., Wang, Z. Q., Li, W. J., Cao, Y. Z., Yu, Y. P., Yu, H., Li, J. H., Ye, J. H., Chen, H., Zhou, Y., Liu, B., Ye, J., Ji, H., Li, S. T., Ni, P. X., Zhang, J. G., Zhang, Y., Zheng, H. K., Mao, B. Y., Wang, W., Ye, C., Li, S. G., Wong, G. K. S. & Yang, H. M. 2004 A Draft Sequence for the Genome of the Domesticated Ailkworm (*Bombyx mori*). *Science* **306**, 1937-1940.

Yu, Z., Ren, M., Wang, Z., Zhang, B., Rong, Y. S., Jiao, R. & Gao, G. 2013 Highly Efficient Genome Modifications Mediated by CRISPR/Cas9 in *Drosophila*. *Genetics* **195**, 289-291.

Zhan, S., Guo, Q., Li, M., Li, M., Li, J., Miao, X. & Huang, Y. 2010 Disruption of an N-Acetyltransferase Gene in the Silkworm Reveals a Novel Role in Pigmentation. *Development* **137**, 4083-4090.

Zhang, J. Z. 2003 Evolution by Gene Duplication: An Update. *Trends in Ecology & Evolution* **18**, 292-298.

Zink, R. M. & Remsen, J. V., Jr. 1986 Evolutionary Processes and Patterns of Geographic Variation in Birds. *Current Ornithology* **4**, 1-69.

Zug, R. & Hammerstein, P. 2012 Still a Host of Hosts for *Wolbachia*: Analysis of Recent Data Suggests That 40% of Terrestrial Arthropod Species Are Infected. *Plos One* **7**, e38544.