The maintenance of genetic variation by balancing selection

Michael Jardine

A dissertation submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

of

University College London

Department of Genetics, Evolution and Environment,
University College London
November, 2021

I, Michael Jardine, confirm that the work presented in this thesis in my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Signature:

November, 2021

Abstract

Adaptive evolution occurs when selection acts on genetic variation for phenotypic traits. In doing so, selection is expected to remove fitness variation in the population. Contrary to this expectation, DNA sequencing has shown that populations harbour high levels of standing genetic variation for fitness. This paradox results in a long-standing question: what maintains genetic variation? One possible mechanism is 'balancing selection', where selection actively maintains polymorphism. Once considered unlikely, studies using genomic and phenotypic approaches have recently given new support for balancing selection and have provided evidence of balancing selection in several species. However, it is often difficult to connect genetic and phenotypic evidence for balancing selection with evidence of the action of selection in real time. This limits our understanding of how balancing selection occurs and its contribution to maintaining genetic variation.

To address these knowledge gaps, I first assayed the fitness effects of a polymorphism in the *Drosophila melanogaster* gene *fruitless*, which shows a signature of balancing selection in wild populations. I show that this polymorphism displays antagonistic pleiotropy, a possible mechanism for balancing selection at this locus (Chapter 2). I next used experimental evolution and pool-sequencing to track the frequency of the *fruitless* polymorphism over time in laboratory populations (Chapter 3). I was able to demonstrate that the *fruitless* polymorphism is probably evolving under balancing selection in these populations, although this result is complicated by 44% of putatively neutral SNPs also being diagnosed as under balancing selection. I next expanded this approach to diagnose selection at 397 candidate sexually antagonistic SNPs. 60% appeared to be under balancing selection (Chapter 4). The equilibrium allele frequency of these SNPs was positively related to that in two wild populations, illustrating that the short-term evolution in the cages is correlated to long-term evolution in wild

populations. That shows that selection is consistent and supports the inference of balancing selection.

Overall, this thesis describes the action of balancing selection in maintaining fitness influencing polymorphisms in *D. melanogaster* and develops methods to diagnose active balancing selection at the population level.

Impact statement

This thesis focuses on balancing selection and its contribution to maintaining genetic variation. Without genetic variation species would be unable to adapt to new environments and speciation would not happen. The forces that maintain genetic variation have long been debated, but recently balancing selection has received revived interest as an important mechanism. However, there are no well described balanced polymorphisms where there is a combination of: a clear genomic signature of balancing selection; the fitness effects of the polymorphism are known; and the polymorphism is shown to be currently evolving under balancing selection. This hinders research into the effects of balancing selection and the role it plays in maintaining genetic variation. To rectify this, I investigated a polymorphism in the *D. melanogaster* gene *fruitless* and show that this both impacts fitness and evolves in a manner consistent with balancing selection. Together with previous evidence, this provides a first example of a well described balanced polymorphism and describes a methodology others could follow to diagnose active balancing selection. I also expanded my analysis to detect balancing selection at candidate sexually antagonistic (SA) sites using a combination of experimental evolution and genome-wide association studies. This method could prove useful to researchers in the future and improve the number of balanced polymorphism loci described.

Outside of academic research, this work has implications for other fields. As climate change and other human impacts continue to threaten species, their genetic variation may be reduced, limiting their ability to adapt to the changing conditions. The maintenance of genetic variation is therefore of great interest to conservation programs which aim to protect species. Genetic variation is also important for agricultural crops which are impacted by climate change and typically harbour low genetic diversity. There is increasing evidence for the role balancing selection plays in maintaining variation in disease resistance loci. Understanding how and why individuals

possess the variation they do could be extremely useful to inform disease responses such as for the covd-19 pandemic. SA loci have also been implicated in the maintenance of several genetic medical conditions and therefore a knowledge of how these types of loci evolve is of great interest to clinicians.

The work described in this thesis has been communicated to the wider scientific community. Data and results contained in this thesis have been presented at the Genetics Society Meeting, 'Genotype to Phenotype to Fitness' (Exeter, 22nd – 23rd November 2018) and at PopGroup 52 (Oxford, 3rd – 6th January 2019). A poster of the finalised work contained in Chapter 2 was presented when I attended the EMBO Course in Population Genomics (Online, 16th – 25th March 2021). Additional presentation of the material presented here has been made at the annual joint DTP conference (London, 12th -13th September 2019). In terms of publication in peer-reviewed scientific journals, the work presented in Chapter 2 was submitted for publication in Proceedings of the Royal Society B, in December 2020, and was published 12th May 2021, (doi.org/10.1098/rspb.2020.2958). I am the first named author on this publication.

Acknowledgements

First and foremost I'd like to thank my primary supervisor Max Reuter. His guidance, patience, and immense scientific know-how have been essential during my PhD and have made me a much better scientist.

Secondly, thanks to my second supervisor Kevin Fowler for his experimental experience, valuable discussions, keeping Max and I on track, and his logistical expertise, particularly in liaising with the DTP and guiding this project through the sometimes labyrinthine UCL admin system.

A big thank you to all those who have contributed to the work in this thesis. Particular mention should go to: Filip Ruzicka without whose efforts much of this work would not exist and for taking me under his wing during my first year; Flo Camus for all her help in the lab and general science advice and who sometimes felt like a third supervisor; Rebecca Finlay whose efforts managing, cleaning, and maintaining the lab were essential for a good research environment; Quentin Saintain who saved me with his expertise for DNA extractions and genotyping; and finally, to Carl Mackintosh who was vital to developing simulations and was patient with a computational novice such as myself.

Thanks to all the others who contributed academically, logistically or socially and who made this PhD a success, including: Raquel Nunes Palmeira, Charlotte Diffley, Mark Hill, Sam Finnegan, Tabitha Owen, Fiona Williamson, Mike Scott, Mike O'Leary, Duncan Greig, Sadé Bates, Josh Schmidt, Enrique Rodriquez, Jacob Maunders, Lara Meade, Will Salvidge, POM, Nick Lane, Nazif Alic, Serian Sumner, Giovanna Vinti, Kieran Samuck, Harvinder Pawar, Paolo Abondio, Olivia Donaldson, Jas Rees, Harrison Ostridge, Marco Colnaghi, Ben Taylor, Emeline Favreau, Chris Brimson, Balint Stewart, Pedram Samani, Dimitri Petrov, Ted Morrow, Jon Harper, and many others.

I’m extremely grateful to my family, Mum, Dad, Granny, Alex, Simon, Lenny, and Ed (RIP), and everyone else who has supported me during this time, I couldn’t have done it without your support.

Also thank to all my friends in London, Scotland and abroad who’ve helped me along the way, especially Tara Copeland, Katrina Grifford, and Rob Baines who have also had to put up with living with me. A shout out to all those from the DTP who made this experience, particularly the first few months so enjoyable, sorry I’ve not been as good a friend to many of you as you have to me.

Finally, a special mention to Jasmine Ono. Your support and encouragement have been invaluable. You’ve picked me up when at my lowest and I wouldn’t have been able to complete this thesis without you. Meeting and getting to know you has been the highlight of my PhD.

Contents

List of Figures

List of Tables

Chapter 1

1. General Introduction

1.1 Genetic diversity

The natural world is incredibly diverse. A wish to account for the vast array of species, their evolution and the various ways they react to their environment, is one of the main goals of evolutionary biology (Darwin 1859; Gloss et al. 2016). This diversity is underpinned at the molecular level by genetic variation. Many genetic variants are neutral, that is they have no impact on the survival or reproduction of the organism in which they are found (Charlesworth and Charlesworth 2010). More important are genetic variants which impact phenotypes, and thereby the fitness of the organism. These fitness affecting variants are the targets of selection (Haldane 1937; Lewontin 1974). Without genetic variation resulting in fitness variation in phenotypic traits, there would be nothing for selection to act on, meaning that no evolutionary change could ever occur (Fisher 1930; Orr 2009). This would prevent species from adapting to meet the challenges to survival such as diseases or environmental change, and new species would never form (Barrett and Schulter 2007; Barton and Keightley 2002). Hence, genetic variation for fitness has been described as the ‘fuel’ which powers selection and evolution (Fisher 1930).

Genetic variation is not only of academic interest to evolutionary biologists, but also important in other fields (Romiguier et al. 2014; Velland and Geber 2005). In conservation, the survival of a species is often dependent on maintaining genetic variation in small and threatened populations so as to preserve their capacity to adapt to changing conditions (DeWoody et al. 2021; Teixeira and Huber 2021). In plants, for example, higher genetic variation is associated with greater establishment success for re-introduced species (Schäfer et al. 2020). Without an adequate volume of standing genetic variation, populations are vulnerable to extinction as adaptation is slow and individuals react in a similar manner to environmental changes (Barret and Schulter 2007). This will become increasingly important as climate change accelerates, forcing species to adapt quickly to new

conditions (Ryding et al. 2021; Pauls et al. 2013). Variation also impacts on the ability of individuals and populations to resist infection by pathogens (Key et al. 2014; DeWoody et al. 2021). In less diverse populations, more members share variation making them more likely susceptible to the same infections (Habel and Schimdt 2012). This is a major issue for threatened species with small population sizes which typically harbour less variation (Hughes et al. 2008; Reed and Frankham 2003). Environmental change and pathogen resistance are of great concern to the agricultural industry as crop strains often have low genetic variation (Pauls et al. 2013; Zhu et al. 2000). For example, bananas almost all originate from a single clonal strain leaving the whole industry vulnerable to new pests due to a lack of immunity (Penna et al. 2019). The introduction of genetic material through breeding and gene editing techniques is currently being investigated to increase the genetic variation in this crop, and others, to protect them from future infections and climate impacts (Penna et al. 2019; Zhu et al. 2000)

While genetic variation is clearly important for selection to occur, it's less clear how variation is maintained (Lewontin 1974). All genetic variation originates as mutations which occur spontaneously in the genome. Most of these mutations are neutral, with no effect on fitness and will evolve purely by drift (Ellegren and Gaultier 2016; Eyre-Walker and Keightley 2007). For those mutations that do affect fitness, directional selection fixes beneficial mutations and removes deleterious ones, thereby reducing the number of variants in the population over time (Lewontin 1974; Kimura 1983; Lande 1976). Thus, variation affecting fitness is not expected to persist for long in the population. However, the first study quantifying genetic variation at the molecular level, Lewontin and Hubby (1966), used allozymes, enzyme variants that differ in structure but not function, to describe genetic variation in wild populations of *Drosophila pseudoobscura*. They found that the level of genetic variation was much higher than previously thought and further studies supported the finding that populations harbour large quantities of genetic variation. But if this is the case, then how is this level of genetic

variation maintained when selection is expected to remove variation? Without a mechanism for it maintenance, genetic variation for fitness won't build up to the levels observed or those required to facilitate the process discussed above. Understanding what maintains this genetic variation is one of the main aims of research in population genetics as we seek to explain why and how life is the way it is (Gloss et al. 2016; Lewontin 1974).

Two main theories for the maintenance of genetic diversity have emerged: (i) mutation-selection balance (Muller 1950), and (ii) balancing selection (Dobzhansky 1955). While both of these phenomena likely occur (Lewontin 1974), the relative contribution of each to the maintenance of genetic diversity for fitness has been fiercely debated for several decades (Beatty 1987). Below I give a brief account of each of these theories and then develop the argument that balancing selection is worthy of further study and investigation.

1.2. How is genetic variation maintained?

1.2.1. Mutation-selection balance

Mutation-selection balance, or mutation-selection-drift balance (MSDB), is one of the major theories to explain how genetic diversity for fitness is maintained (Leffler et al. 2012; Mitchell-Olds et al. 2007). MSDB describes a situation where the level of genetic variation in the genome is determined by the balance of mutation creating new variation and, on the other hand, selection and drift removing variation (Kimura and Ohta 1971). Considering a single 'average' locus in a diploid organism, new mutations are created at a rate, μ. They are removed from the population by selection, with strength s, with the effectiveness of selection dependent on how visible the mutation is to selection, i.e. its dominance, h (Crow and Kimura 1970). The effects of mutation and selection reach an equilibrium where the expected frequency of a mutation, q, in the population occurs where the rate at which new mutations occur equals the rate at which selection removes them, $q = \mu/hs$

(in the absence of drift) (Crow and Kimura 1970). It is easy to see that the equilibrium frequency (and hence genetic variation) will vary depending on the values of the parameters. Increased mutation rate will lead to higher equilibrium frequencies, while increased strength of selection and dominance will reduce *q*.

MSDB was a popular theory for the maintenance of genetic variation for fitness, in part due to its simplicity (Barton and Keightley 2002). Drift also plays an important role through the random selection of alleles at each generation which can end up fixing or removing variants, especially in smaller populations (Lynch 2011; Kimura and Ohta 1971). Population size, and in particular effective population size, is therefore an important factor governing the amount of genetic variation maintained under MSDB as larger populations are less vulnerable to the effects of drift and selection against deleterious mutations is more effective (Leffler et al. 2012; Ellegren and Gaultier 2016).

Mutation-selection balance has been shown to be theoretically capable of maintaining genetic diversity in populations (Lande 1975; Desai and Fisher 2007). Assuming that most novel mutations are mildly deleterious and recessive, MSDB predicts that most variants will exist at low frequencies as selection is less effective at removing recessive mutations or those at very low frequencies (Kimura and Crow 1964; Crow and Kimura 1970). While evidence for such patterns has been found in *D. melanogaster* (Elyashiv et al. 2016), MSDB is not capable of explaining patterns of genetic diversity where polymorphisms are maintained at intermediate frequencies or for long-lived polymorphisms (Charlesworth 2015; Fijarcsk and Babik 2015) as have been observed in genetic data. By comparing measures of genetic variation from multiple studies of *D. melanogaster,* Charlesworth (2015) proposed that although MSDB does contribute to genetic variation for fitness, it cannot account for all the variation observed in this species. Coupled with increasing evidence for balancing selection detected by genomic and phenotypic

studies from both *D. melanogaster* and other species (Lindtke et al. 2017; Mérot et al. 2020), this means that genetic diversity being maintained solely by MSDB is not the full story. This represents a shift in understanding from past decades, and now the main question is what other mechanisms could account for the disparity between the observed and explained volume of standing genetic diversity? The major alternative theory is balancing selection.

1.2.2. Balancing selection

Balancing selection describes a situation where no one allele at a locus confers a consistent fitness advantage (Gloss and Whiteman 2016). That is, no one allele results in producing a phenotype that is always favoured by selection (Llaurens et al. 2017). For example, consider a locus with two alternative alleles: A and B. Within one selective environment, allele A is the fitter allele and individuals carrying this allele are selected for. The frequency of A in the population thereby increases. However, in a different selective environment the situation flips and now the B allele is fittest, resulting in an increase in the frequency of the B allele. In cases where both the environment favouring A and the environment favouring B are not consistent over space and time, neither allele reaches fixation, as neither is always fitter than the other. This inconsistency on selective pressures which maintains genetic variation at this locus is called balancing selection (Charlesworth 2015, Lewontin 1974) (Figure 1.1).

The selective environment is any intrinsic or external condition where an allele may find itself and which affects its fitness selection. This could be external climatic conditions, the life-history stage of the organism the allele is present in, the organism's sex, or the genotype of other population members. Different mechanisms to describe the process by which balancing selection occurs have been proposed. These mechanisms include: negative frequency dependent selection (Sinervo and Lively 1996; Amambua-Ngua et al. 2012), reciprocal sign epistasis (Ono et al. 2017), heterozygote advantage

(Johnstone et al. 2013), fluctuating environmental conditions (Abdul-Rahman et al. 2021; Bergland et al. 2014), sexual antagonism (SA) (Pennell and Morrow 2013), and antagonistic pleiotropy (AP) (Rose 1982). Despite the number of alternative mechanisms, the general principle of action is roughly the same, namely that selection is context dependent (Charlesworth and Hughes 2000; Llaurens et al. 2017).

Balancing selection is an intriguing solution to the problem of what maintains genetic variation. This theory is now receiving the recognition that it deserves and research seeking to identify loci under balancing selection and their role in evolution is increasingly common (Fijarcsk and Babik 2015). However, why was balancing selection long considered to contribute little to the maintenance of genetic in comparison to MSDB, and what developments have occurred to change this situation?

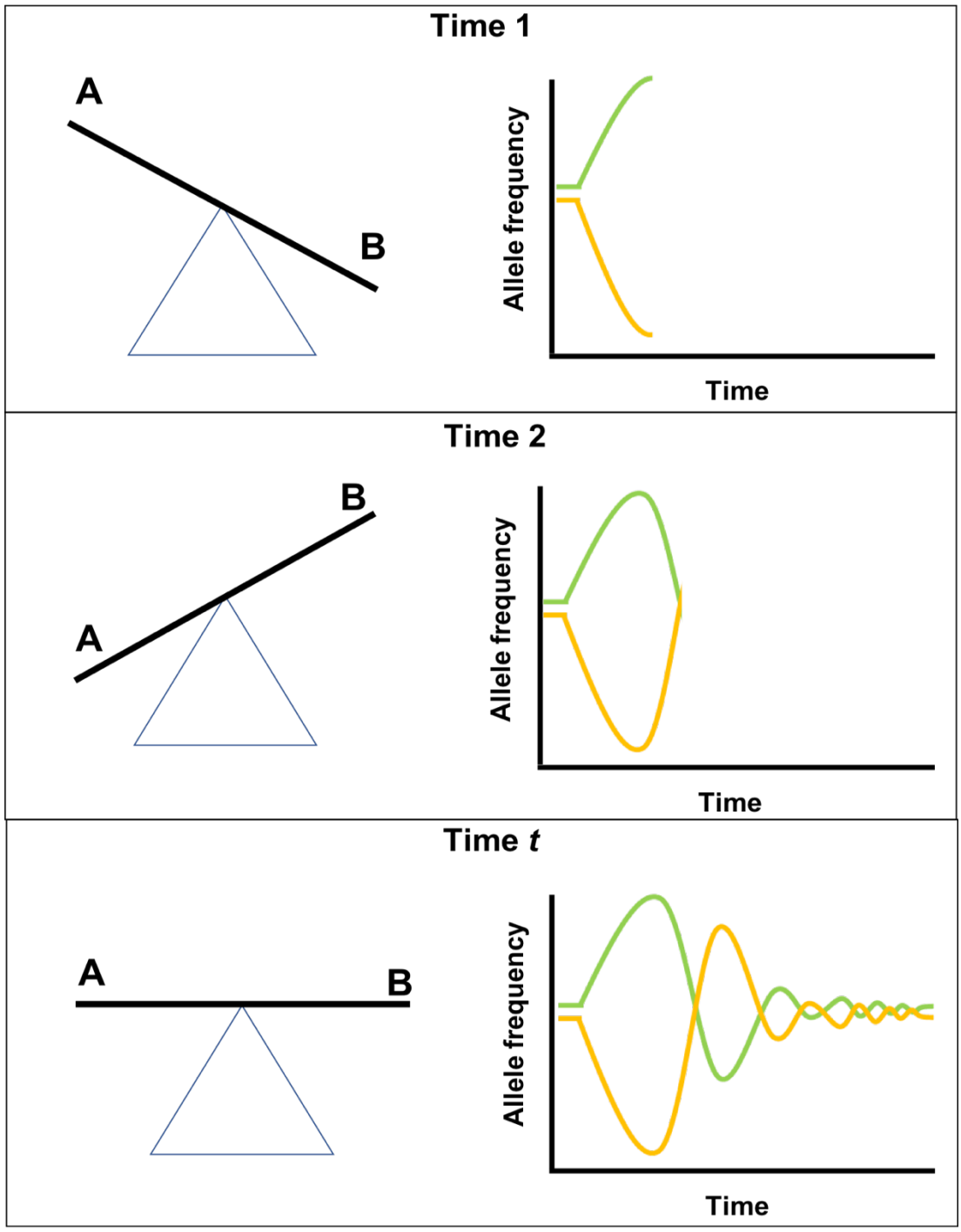


Figure 1.1. Balancing selection and its effect on allele frequency. Assume a population with a locus polymorphic for two alleles A and B. Both alleles are present in roughly equal numbers in the population. At Time 1 some perturbation (such as a seasonal change in temperature) occurs causing the fitness of allele A to be greater than that of B. The frequency of A therefore increases. At a later point, Time 2, this perturbation shifts and now allele B is fitter and the allele frequencies return towards

their starting positions. After Time *t* of subsequent perturbations or generations, the fitness of the 2 alleles balance out and the frequency of the two alleles reach an equilibrium point. Both alleles are retained and genetic diversity for fitness is maintained at this locus through balancing selection. This is a theoretical case and specific mechanisms of balancing selection will produce different patterns. Loci affected by seasonal fluctuations in selection will not reach an equilibrium point, but will cycle over long time periods. However, for loci affected by antagonistic pleiotropy, it would be very difficult to detect any fluctuation in allele frequencies.

1.3. The case for balancing selection

1.3.1 Initial support and criticism of balancing selection

Balancing selection as a means to maintain genetic diversity for fitness was initially suggested by Theodosius Dobzhansky in the 1950s (Dobzhansky 1955). Building on the theoretical work of Levene (1953) and the investigation of sickle-cell disease (Haldane 1949; Allison 1954), Dobzhansky proposed that fluctuations in the strength and direction of selection could potentially maintain genetic variation (Gloss and Whiteman 2016). The balancing selection view postulates that alternative alleles could be maintained at intermediate frequencies, and for long periods of time, and that variation would have some fitness benefits (Dobzhansky 1970; Charlesworth and Huges 2000). This clashed with the ‘classical’ mutation-selection balance model (Beatty 1987; Muller 1950). Since this was in the days before genetic variation could be measured at the molecular level, the argument for balancing selection was primarily theoretical (Beatty 1987; Charlesworth 2015). This led to a number of theoretical models being developed to investigate how the strength, timing, population size, fluctuation frequency, and various other selection parameters impact the probability of balancing selection occurring (Clarke 1972, 1979, Dempster 1955, Haldane and Jayakar 1963, Levene 1953, Hedrick 1976, 1986, 1990, Gillespie and Turelli 1989, Turelli and Barton 2004, Lande and Shannon 1996; Kidwell et al. 1977; Felsenstein 1976). Although many of these have been criticised for lacking biological realism and robustness since they largely only considered

alleles of large affect in haploid organisms with narrow niches (Maynard Smith and Hoekstra 1980; Hoekstra et al. 1985), they did demonstrate that, as long as each alternative allele is favoured by selection at one time or another, balancing selection could develop and maintain genetic variation (Gloss and Whiteman 2016; Llaurens et al. 2017).

The first molecular investigations of genetic variation emerged in the 1960s with Lewontin and Hubby's work using allozyme electrophoresis to provide a measure of genetic diversity in *Drosophila pseudoobscura* (Lewontin and Hubby 1966; Hubby and Lewontin 1966). This work initially seemed to support the picture proposed by balancing selection, that genetic variation was abundant in natural populations, and launched a host of similar studies assaying allozyme variation (Charlesworth 2010; Lewontin 1974). However, two important factors led to a decline in the belief that balancing selection played a major role in maintaining genetic variation.

The first was that although allozyme data showed that variation was abundant, there was little evidence that this variation had any effect on fitness (Charlesworth 2010; Casillas and Barbadilla 2017). Such variation could therefore be neutral and play no real role in evolution, so could not be maintained by balancing selection. This gave credence to the 'neutral theory' of molecular evolution (Kimura 1968) which proposed that most fitness affecting mutations will be deleterious and selected against. Therefore, the majority of the genetic variation observed is due to neutral mutations evolving by drift (Kimura 1968). Variation could also exist temporally by rare beneficial mutations moving towards fixation, known as a selective sweep (Maynard Smith and Haigh 1974). This theory, and the later developed 'nearly neutral theory' (Ohta 1973) were compatible with the observed patterns from allozyme data and supported MSDB being the major process by which genetic variation was maintained (Otha 1992; Mitchell-Olds et al. 2007).

The second factor which led to a decline in support for balancing selection was that although several models had been proposed describing the conditions under which balancing selection could evolve, actual examples were difficult to find (Charlesworth and Hughes 2000; Prout 2000). In the absence of a clear link between molecular markers and fitness effects, the demonstration that traits and their underlying molecular variation were under balancing selection was very difficult to prove (Gloss and Whiteman 2016). This compared unfavourably with the combination of the neutral theory and MSDB which were more compatible with the existing molecular and theoretical evidence at the time (Casillas and Barbadilla; Charlesworth 2010). The role of balancing selection was considered to play only a minor role, reserved for a few cases where clear-cut examples of heterozygote advantage were described (Mitchell-Olds et al. 2007; Leffler et al. 2012). Such examples included major genetic diseases in humans such as sickle-cell anaemia (Haldane 1949) and cystic fibrosis (Gabriel et al. 1994), or the self-incompatibility loci of plants (Wright 1939).

Without clear examples of balanced polymorphisms, balancing selection theory also struggles to account for the detrimental effects of genetic load (Bernatchez 2016). The volume of less fit or deleterious mutations which the genome carries, referred to as load, negatively impacts the average fitness of individuals within the population and hinders the process of adaptation (Kimura 1968, Agrawal and Whitlock 2012). This is a problem with balancing selection which maintains more genetic variation than MSDB. Since balanced variation is maintained due to beneficial effects under certain circumstances, the rest of the time it may be detrimental, or at least not as fit as other alternative variation (Haldane 1957). Hitchhiking of linked deleterious mutations to loci under balancing selection can also contribute to the volume of load (Kimura 1968; Llaurens et al. 2017). This hitchhiking variation can be a bigger problem as it partially offsets the beneficial effects of balanced variation. An additional problem, referred to as segregation load, occurs for cases of heterozygote advantage, as sexual reproduction creates

a number of unfit homozygotes at each generation (Uyenoyama 2005). Kimura (1968) pointed out that the volume of genetic load that would be incurred if all of the genetic diversity measured from allozyme data was due to balancing selection would place too great of a strain on the population. In light of these various factors, the proposition that balancing selection was a major contributing mechanism to maintaining genetic variation became more difficult to support (Lewontin 1974).

1.3.2 The renaissance of balancing selection

However, things has changed somewhat over the past couple of decades, forcing us to take a closer look at the potential contribution of balancing selection towards maintaining genetic variation (Hedrick 2006). This has occurred in part due to the rapid development of sequencing technology. Modern next-generation technology is able to produce sequencing data of a volume many orders of magnitude greater than that which was possible in the past (Casillas and Barbadilla 2017). These developments mean that high quality whole genome sequences can be rapidly and cheaply generated for individuals or whole populations. Sequence polymorphisms can be accurately linked to real-life fitness effects – something that is not so easy with allozyme data (Charlesworth 2010). This has provided the opportunity to investigate how the genome responds to selection at the molecular level (Casillas and Barbadilla 2017). Modern sequencing technology allows the tracking of allele frequencies in real time (Franssen et al. 2017) while genomic scans allow us to search the genome for signatures of variation expected to be generated by balancing selection (Andrés et al. 2009, Bitarello et al. 2018; Leffler et al. 2013; Karasov et al. 2014). Sequencing can also uncover the genetic architecture of phenotypic traits which were previously not possible (Llaurens et al. 2017). One important conclusion from new sequencing was made by Charlesworth (2015). As mentioned in section 1.2.1., Charlesworth compared studies measuring the amount of genetic variation for fitness in *D. melanogaster*. He showed that the amount of genetic diversity for fitness cannot be explained by MSDB alone. The

difference must be made up by variants maintained by balancing selection (Charlesworth 2015).

Secondly, there has been a shift in focus away from primarily considering balancing selection in the context of overdominance, and other mechanisms capable of generating balancing selection have gained greater attention (Hedrick 2012; Novak and Barton 2017). One such class of polymorphisms are antagonistic loci which generate balancing selection not as a result of their zygotic state, but because each one is fitter under different circumstances resulting in conflict between two alternative alleles. These types of polymorphisms typically don't suffer from the issue of segregation load in the same way as those expressing overdominance do (Haldane 1957; Kimura 1983). Several studies have found evidence for antagonistic variation in nature including in *D. melanogaster* (Rice 1992), bank voles (*Myodes glareolus*) (Mokkonen et al. 2011), and red deer (*Cervus elephus*) (Forester et al. 2007). These examples all describe sexually antagonistic variation, a situation where variation at a locus affects a trait in both males and females, but where the sexes disagree on which is the fitter allele (Bonduriansky and Chenoweth 2009). This is a potential source of balancing selection as fitness variation is maintained due to conflict between the sexes. This type of genetic variation may be common given that the sexes often differ in the ways they can maximise their fitness (Connallon and Clark 2014a; Pennell and Morrow 2013). Other mechanisms of balancing selection, such as antagonistic pleiotropy, where the fittest variant varies between life-history stages (Williams 1957), have received less attention, possibly a result of theoretical studies doubting the role such loci play in maintaining variation (Curtsinger et al. 1994; Hedrick 1999). However, recent theoretical (Zajikteck and Connallon 2018) and empirical work (Mérot et al. 2020) show that the dismissal of antagonistic pleiotropy as an important mechanism for maintenance of variation may be premature. Consideration of other types of polymorphisms increases the opportunity that variation and fitness patterns resembling balancing selection will be described. New instances generate

enthusiasm for the field, further increasing our ability to examine the role of balancing selection as a viable explanation for maintaining genetic variation for fitness.

Thirdly, theoretical developments have suggested that the genetic load generated by balancing selection could be mitigated by epistasis, dominance reversal, differential expression, or by fixation in small populations (Connallon and Chenoweth 2019; Agrawal and Whitlock 2012). A number of the classic examples of balanced polymorphism such as sickle-cell anaemia are Mendelian traits, where one locus is responsible for the phenotype. However, many traits are polygenic in nature where the effect each locus has on fitness is small (Lande 1981; Wright 1968) Therefore, the negative impact of an ill-fitted allele will not have such an acute disadvantage on an organism's fitness since it only has a small impact on fitness overall (Yeaman 2015; Yeaman and Otto 2011). The load associated with each locus will also be reduced as more time is available for recombination to break up associations between a balanced polymorphism and load before negative impacts are felt (Barton and Keightley 2002; Kardos and Luikart 2021). Other work has shown that the negative consequences of balancing selection are lessened, and balancing selection is more likely to establish, when mutations with varying effect sizes, and direction, on fitness are considered. For example, Sellis et al. (2011) showed that, when considering balancing selection in the framework of Fisher's geometric model, balanced polymorphisms could stabilise variation during a trait's adaptive walk towards its optima, and provide an insurance against changing conditions. The potential for balanced polymorphisms to become established is even greater for dioecious species as sex-dependant selection increases the number of means by which variable selection may occur (Connallon and Clark 2014b). Different forms of balancing selection may also influence each other, facilitating the establishment of multiple mechanisms to generate balancing selection (Zajikteck and Connallon 2018). Genotype by environment interactions may also increase the probability that balancing develops and

can maintain genetic variation, especially for polygenic traits (Turelli and Barton 2004). Early models of balancing selection were criticised for lacking biological realism (Maynard Smith and Hoekstra 1980) which contributed to balancing selection being difficult to justify. However, many recent models include greater biological realism and also support the idea that balancing selection can evolve and maintain genetic variation for fitness (Wittman et al. 2017; Tellier et al. 2007).

In light of these various developments, the plausibility of balancing selection as a major contributor to maintaining genetic variation for fitness has been reassessed. The detection and identification of loci under balancing selection and how these contribute towards the maintenance of genetic variation for fitness is not only of great interest to resolving a key issue in population genetics. By maintaining genetic variation for fitness, balancing selection has the potential to have large impacts on aspects of health care, conservation, and agriculture (Amambua-Ngua et al. 2012; Barret and Schulter 2007; Reed and Frankham 2003). To understand the contribution and mechanisms by which balancing selection occurs, it is necessary to both identify loci under balancing selection and gain some insight into the phenotypic effects of balanced polymorphisms. To this end, two avenues of investigation have proved fruitful: (i) detection of balancing selection from genomic data, (ii) studies relating genotypic data to fitness measures. I will now review some of the findings from these investigations and explain how they contribute to the understanding of the maintenance of genetic diversity by balancing selection.

1.3.3 Evidence for balancing selection from genomic data

The patterns of genetic variation are expected to vary depending on the mode of selection acting (Charlesworth 2006; Charlesworth et al. 1997). Directional selection reduces genetic variation through fixation or removal of mutations, not just at the site under selection but also at linked sites (Charlesworth 1993). Neutral variation will fluctuate over time due to drift in

accordance with the population size (Charlesworth et al. 1995; Kimura 1968). Balancing selection on the other hand is expected to result in more elevated and consistent levels of nucleotide variation as alternative alleles can contribute to fitness (Fijarcsk and Babik 2015). As with directional selection, linkage maintains variation not just at the site under selection, but at linked neutral loci (Charlesworth 2006). This creates a distinct allele frequency spectrum indicating greater variation around a balanced polymorphism (Charlesworth 2006; Fijarcsk and Babik 2015). Therefore, the genome can be scanned for regions where the allele frequency spectrum is indicative of balancing selection rather than neutrality (Andrés et al. 2009; Charlesworth 2006; Fijarcsk and Babik 2015). This search often makes use of statistics such as Tajima's D where values > 0 can indicate balancing selection, and new methods are routinely developed to improve the process (Bitarello et al. 2018; Isildak et al. 2021; Siewert and Voight 2017). Long-term maintenance of genetic variation is another classic signature of balancing selection (Fijarcsk and Babik 2015). This can be detected using coalescence time, an estimate of how long it has been since sequences shared a common ancestor (Takahata and Nei 1990). Regions under balancing selection will have longer coalescent times than neutral ones (Fijarcsk and Babik 2015; Richman 2000).

Signatures of long-term balancing selection have been detected for several genes in humans. By comparing genome sequences from 39 people from two distinct populations, (20 of European decent and 19 of African) Andrés et al. (2009) identified 60 genes with a strong signature of balancing selection based on increased estimated coalescence times and an excess number of intermediate frequency alleles. These genes have likely been under balancing selection for thousands of years, at least since before the split between the two populations. While 60 genes may not be a large number considering the thousands of genes in the human genome, it nevertheless represents an important finding as even a single gene can have wide-ranging effects on fitness and therefore genetic variation. The

identification of specific genes, such as immunity related genes, can be informative about the types of genes that may be targets of balancing selection (Andrés et al. 2009). As the number of genomes that have been sequenced has increased and detection methods have improved so has the rate of detection of balancing selection patterns. A recent study demonstrated the use of a new statistic called non-central deviation (NCD). Using this technique to scan for signatures of balancing selection in genomes of four human populations from the 1000 genomes project (1000 genomes project 2012), Bitarello et al. (2018) found evidence for balancing selection covering 1.6% of the human genome. This included 1,594 genes (8.5% of all protein coding genes), meaning that a sizable proportion of human variation for fitness could be under balancing selection.

Long-term balancing selection can even extend over millions of years and be retained after speciation events (Fijarcsk and Babik 2015; Richman 2000). Signatures of balancing selection have been found to be shared between humans and chimpanzees in immunity genes and in another 125 regions where the same haplotypes were segregating in the two species (Leffler et al. 2013), and also potentially between humans and Neanderthals (Hsieh et al. 2021), demonstrating that balancing selection can maintain genetic diversity over very long periods of time. This first study in chimpanzees (Leffler et al. 2013) is an example of the progress that can be made thanks to modern sequencing technology; a similar study conducted 8 years before using an earlier, less complete Chimpanzee genome, failed to find any evidence of shared balancing selection (Asthana et al. 2005). As genome sequencing develops, and more nuanced methods of detecting selection follow, our ability to detect signatures of balancing selection will likely become more sensitive and refined.

Genomic scans have frequently detected balancing selection at loci associated with evolutionary arms races, particularly those involved in immunity and pathogen response (Key et al. 2014). In humans, genes which

form part of the major histocompatibility complex (MHC or the Human Leukocyte Antigen system—HLA) are responsible for much of our immune response (Harding and Geuze 1993). These genes are locked in an arms race with various pathogens as rarer alleles are less likely susceptible to attack, and pathogens adapt to our defences and will tend to target common genotypes (Ejsmond and Radwan 2015). This causes negative-frequency dependent selection on both our MHC genes and the invasion genes of pathogens which can lead to balancing selection (Key et al. 2014; Hedrick 1998; Ejsmond and Radwan 2015; MacPherson et al. 2020). MHC genes display elevated levels of genetic diversity consistent with balancing selection (Key et al. 2014). For example, the peptide loading genes ERAP1 and ERAP2 are associated with defence against viral pathogens including HIV (Andrés et al. 2010; Cagliani et al. 2010). MHC genes are not just present in humans but in all vertebrates and balancing selection at these loci are important in maintaining diversity in pathogen resistance in many species (van Oosterhout et al. 2006). Selection at the MHC also results in corresponding selection at invasion loci in parasites and identification of such loci is an important tool in tackling infectious diseases, focussing research and developing prophylaxes or cures (Amambua-Ngua et al. 2012).

The role of balancing selection maintaining genetic diversity at pathogen resistance loci appears widespread and has been detected at both resistance genes in plants (Höger et al. 2012; Karasov et al. 2014) and antimicrobial peptide (AMP) genes in *Drosophila* (Unkless et al. 2016). AMPs are highly variable in dipteran species and appear to have been retained across speciation events (Unkless et al. 2016; Chapman et al. 2019). Medically important genes under balancing selection are not limited to those responsible for immunity. Sato and Kawata (2018) detected balancing selection at the mammalian gene *SLC18A1* in humans. The variant *Thr136Ile* has been associated with bipolar disorder and anxiety personality traits (Lohoff et al. 2006, 2008), but is maintained at intermediate frequencies of between 20-61% in non-African populations (Sato and Kawata 2018). Quite

why this gene is under balancing selection is unclear, but it does contribute to understanding how variation related to serious psychological disorders persists.

Although pathogen resistance and immunity genes in humans and *Drosophila* have been the most commonly identified targets of balancing selection, other studies have found signatures of balancing selection for loci associated with sexual conflict (Sayadi et al. 2019), sex determination (Ding et al. 2021), and colour morphology (Lindtke et al. 2017) in a range of organisms. Croze et al. (2017) identified 183 candidate genes under balancing selection in *D. melanogaster* which included a richness of genes involved in neural development and circadian rhythm, although the strongest signals were found for immunity genes. The increasing ease of collecting genomic data from patients has facilitated the construction of large genomic datasets such as the UK's 'bio-bank' (Bycroft et al. 2018). These datasets are a valuable resource and allow the detection of genetic variants under balancing selection, not just at sites associated with immunity to pathogens, but also those maintained by other mechanisms such as sexual antagonism (Ruzicka et al. 2021).

An issue with studies which detect balancing selection using genomic approaches is that they work by comparing the observed allele frequency spectrum to an ideal distribution where loci are evolving either neutrally or under directional selection (Fijarcsk and Babik 2015). Therefore, the chances of detecting balanced polymorphisms are greater the longer polymorphisms have existed, where demographic changes have been small, and where equilibrium allele frequencies are nearer to 0.5, as the different selective forces have had a greater chance to become established (Charlesworth 2006; Fijarcsk and Babik 2015; Charlesworth et al. 2003; Clark 1997). More recent or short-lived balanced polymorphisms will be missed as they produce less distinct genomic signatures (Charlesworth et al. 1997). Such cases are therefore difficult to distinguish from soft selective sweeps or neutral variation

under drift (Fijarcsk and Babik 2015; Hermission and Pennings 2005). Other and more sensitive methods are required to try and detect such balanced polymorphisms.

One such approach used to detect loci under selection has been evolve and re-sequence studies (E&R) (Schlötterer et al. 2015, Kofler and Schlötterer 2014; Garland and Rose 2009). This is a form of experimental evolution (Kawecki et al. 2012) where an experimental population is maintained under a set of selective conditions, for example, a warmer environment than they are used to (Orozco-terWengel et al. 2012), and the population's evolutionary response is measured (Schlötterer et al. 2015). The outcomes are a consequence of the conditions which are applied to the whole population. Such studies could be considered a form of 'laboratory natural selection' rather than being a 'true' experiment. In contrast, with artificial selection the individuals contributing to the next generation are specifically chosen based on some trait value which enforces a correlation between that trait and fitness (Kawecki et al. 2012; Garland and Rose 2009). In E&R studies, after a number of generations, a sample of individuals is taken from the experimental population and sequenced. By comparing changes in allele frequency between the beginning and the end of the experiment, it is possible to detect which loci are under selection during the experiment (Schlötterer et al. 2015).

Although E&R studies have not been widely applied to detect balancing selection, they have been successful in identifying genes under directional selection in *D. melanogaster*. Different studies have identified genes responsible for adaptation in temperature tolerance (Orozco-terWengel et al. 2012; Tobler et al. 2013); pathogen and disease resistance (Jalvingh et al. 2014; Martins et al. 2014); and lifespan (Remolina et al. 2012). Signatures of selection can be difficult to detect and vary between experiments meaning that usually results are compared to simulated datasets to identify the loci involved (Franssen et al. 2017; Kofler and

Schlötterer 2014). However, it can be difficult to confirm that an allele frequency change at a particular locus is actually contributing to adaptation and is not due to other factors such as selection at linked loci (Kapun et al. 2014; Franssen et al. 2015). It is sometimes also difficult to understand how changes in variation manifest phenotypically to create fitness differences acted on by selection. The E&R approach has not yet been widely applied to detect loci under balancing selection although there are some recent examples (Chelo and Teotónio 2013; Kazangıoğlu and Arnqvist 2014). One recent study used an E&R methodology to track mitochondrial haplotypes in *D. subobscura* (Novičić et al. 2020). They found evidence of negative-frequency dependent selection which maintained two genetically diverged mitochondrial haplotypes in the population, likely through differing effects on resource acquisition (Novičić et al. 2020). A second study established 10 field populations of *D. melanogaster* to track how environmental fluctuations affect selection over the course of the year (Rudman et al. 2021). They found that selection fluctuated over time resulting in adaptive allele frequency shifts at many independent loci which contributed to maintained genetic variation at these loci.

A related approach of sequencing a population at multiple time points has been performed for wild *D. melanogaster* populations. Bergland et al. (2014) discovered that allele frequencies varied throughout the year but were consistent between years. Modelling supported that this was due to seasonal fluctuations in environmental conditions over the year (Wittman et al. 2017), and such patterns were consistent with balancing selection maintaining diversity for seasonally adapted traits. While the precise effect on fitness provided by individual fluctuating polymorphisms is unclear, such patterns have been observed in *D. melanogaster* populations across different latitudes (Machado et al. 2021) and continents (Kapun et al. 2020). A similar result was found in the mussel species *Septifer virgatus*, where genetic variation associated with thermal tolerance covaried throughout the year with the ambient temperature of the water (Han et al. 2020). Such temporally

modulated variation could have serious knock-on effects for populations as climate change destabilises climate patterns (Friedman et al. 2017; Ryding et al. 2021). This could result in a halt to balancing selection helping to maintain the genetic variation in some species leaving them more vulnerable to environmental change (Han et al. 2020). On the other hand, species which experience fluctuating environmental selection may possess higher genetic variation allowing for greater adaptive potential to changing conditions (Barrett and Schulter 2007; Friedman et al. 2017; Wittman et al. 2017). Either way, the sequencing of populations has shown that balancing selection has an important part to play in the survival of species in a changing world.

1.3.4 Phenotypic examples of balancing selection

An alternative approach to studying balancing selection comes from identifying instances where variation in phenotypic traits results in corresponding variation in fitness. Cases where phenotypic variation is maintained over time due to different phenotypes having advantages over others at different times, or where variation is maintained despite an apparent directional relationship between the phenotype and fitness, are likely to be under balancing selection. Variation in the fitness of these phenotypes translates to the molecular level and maintains genetic variation for fitness. Identifying such traits can be a difficult since the recording of the necessary fitness information is a laborious task taking much time and effort (Pardo-Diaz et al. 2015; Wilkinson et al. 2015; Mullon 2012). However, several studies have identified phenotypic variation in fitness related traits that is maintained by balancing selection.

One classic example comes from side-blotched lizards (*Uta stansburiana*) (Sinervo and Lively 1996). Males of these species have three different colour morphs, orange, blue and yellow, each of which displays a specific mating strategy which are respectively: aggressive ‘ultradominance’, mate guarding, and sneaking (Sinervo et al. 2000). Each morph outcompetes one of the other morphs, but loses out to another in rock-paper-scissors type

dynamics (Maynard Smith 1982) where each morph and behavioural strategy is maintained (Sinervo and Lively 1996). The fittest phenotype depends on which other phenotype is most common at any one time, an example of negative frequency-dependent selection (Allen and Clarke 1982). While this clearly demonstrates an ability to maintain phenotypic diversity and therefore an example of balancing selection in action, the genetic determinants of this trait are unknown. However, it has been found that populations of side-blotched lizards have particularly high levels of genetic diversity (Micheletti et al. 2012). It remains to be seen if this diversity is truly a result of the balancing selection on male mating strategy, but alternative reproductive traits such as these are a prime candidate for balancing selection (Stewart et al. 2019).

Another study carried out on a population of Red Deer (*Cervus elaphus*) on the Isle of Rum recorded the life-time reproductive success of individuals to form a genealogy of the populations over several generations (Foerster et al. 2007). This was analysed using an 'animal model' to estimate a value of life-time reproductive success for each individual (Kruuk 2004). They found that males with higher than average lifetime reproductive success produced daughters with relatively low fitness (Foerster et al. 2007). This points towards the presence of sexually antagonistic (SA) genetic variation, that is, variation that contributes to high fitness in one sex but poor fitness in the other (Cox and Calsbeek 2009; Bonduriansky and Chenoweth 2009). SA relationships can generate balancing selection as variants are selected for in one sex but not in the other and thereby contribute to the maintenance of genetic variation (Connallon and Clark 2012; Mullon et al. 2012). However, without genomic data we are not able to identify the specific genes responsible for this SA relationship and for maintaining variation for lifetime reproductive success.

Other studies have documented balancing selection on individual genes. For example, variation in male horn size of wild Soay Sheep (*Ovis*

aries) is mostly controlled by a single gene, RXFP2 (Johnstone et al. 2011). This gene has two alternate alleles. Ho^{+} confers greater horn size and thus greater reproductive success, while the Ho^{P} allele grants greater survival to adulthood (Johnstone et al. 2013). Those males heterozygous at this gene have greater lifetime fitness than either homozygote, therefore it appears that genetic variation for this trait is maintained by heterozygote advantage (Johnstone et al. 2013). Both this example and that of the Red Deer cited above, come from long-term closely studied populations of animals on isolated islands and these aspects may contribute to the identification of genetically derived fitness traits. These types of polymorphism are likely common in other species where populations are not so well studied or isolated.

Sometimes a change in environmental pressures can reveal genes under balancing selection by altering how some traits are selected. One example is the DDT resistance allele, DDT-R, of the gene *Cyp6g1* in *D. melanogaster*. This allele was initially intermediate in populations until the widescale spraying of DDT as an insecticide (Daborn et al. 2002). After this there is evidence of a strong selective sweep in a 20Kbp region around the *Cyp6g1* gene (Daborn et al. 2002). Apart from the fitness advantages when DDT is present, DDT-R females also show greater fitness in the absence of DDT. They lay more eggs with greater viability than do non DDT-R females (McCart et al. 2005). Why therefore did the DDT-R allele not sweep through the population before the addition of pesticide? Further investigation found that although DDT-R was beneficial in females it had adverse effects on fitness in males resulting in smaller males with lower mating success (Smith et al. 2011). It was proposed that variation at the *Cyp6g1* gene was maintained due to SA effects on the mating performance of the two sexes. But after spraying of DDT this balance broke down and the DDT-R allele swept to fixation, decreasing the genetic diversity in this region (Smith et al. 2011). The strength and duration of balancing selection is sensitive to influence from external factors in the environment and changes in selection

regime experienced by linked loci (Turelli and Barton 2004). Genomic investigations are effective for long-term detection of balancing selection, but struggle to identify more temporal short-term balancing selection since genomic signatures can take time for detection to be possible (Fijarcsk and Babik 2015). In the case of the *Cyp6g1* gene the disruption of balancing selection resulting in a loss of genetic diversity would have been difficult to uncover using genomic inferences alone.

Ultimately a combination of approaches is needed to identify loci under balancing selection and to quantity their phenotypic effects. A recent study employed various methods to identify the variation responsible for life-history differences in the seaweed fly *Coelopa frigida* (Mérot et al. 2020). This species contains a large inversion with two allelic forms: α and β (Mérot et al. 2018, Day et al. 1983). The allele affects the size of the adult fly, αα flies are larger than αβ flies, which are in turn larger than ββ flies (Butlin and Day 1984, 1985). This size difference has knock-on effects for reproduction where larger flies are generally fitter in both sexes, but the frequency of the two alleles are roughly equal in wild populations. Mérot et al. found that although αα males achieved more matings and αα females laid more eggs, both sexes had poorer survival to adulthood and αα males also took nearly twice as long to develop than ββ individuals (Mérot et al. 2020). Heterozygotes had intermediate performance compared to homozygotes for both reproduction and development. This relationship where each allele is fitter than the other at a different life-history stage is an example of antagonistic pleiotropy (Rose 1982, 1985). Using a series of simulations Mérot et al. showed that the differences in fitness were substantial enough to create balancing selection across the inversion as the seaweed substrate on which the flies live is only available for a short and variable amount of time. The two alleles differ by >2.5% in the coding regions (Mérot et al. 2018), thus this polymorphism is an example where both the genotype and the environment interact to maintaining genetic diversity through balancing selection (Turelli and Barton 2004). The identity of the causative variant as

an inversion is interesting and points towards the potentially important role that inversions play in maintaining genetic diversity via balancing selection and adaptation more generally (Mérot et al. 2021; Wang et al. 2013; Kapun and Flatt 2019).

Another study used a combination of phenotypic and genomic approaches to identify balancing selection maintaining a regulatory polymorphism in the *D. melanogaster* X-linked gene *fezzik* (*fiz*) (Glaser-Schmitt et al. 2021). A single SNP located in a regulatory region of the *fiz* gene had previously been identified in all non-African populations of *D. melanogaster* at roughly intermediate frequencies (Glaser-Schmitt and Parsch 2018; Saminadin-Peter et al. 2012). Through repeated annual sampling of a wild population Glaser-Schmitt et al. (2021) discovered that the frequency of the *fiz* polymorphism fluctuated depending on the season in females, but not males, and the overall frequencies were consistent year-on-year, similar to results recorded by Bergland et al. (2014). Modelling showed that the frequency dynamics at this locus were consistent with the action of both environmentally fluctuating selection and sexual antagonism, both sources of balancing selection. Expression patterns found that the two *fiz* alleles varied in their phenotypic dominance, and that this was dependent on their life-history stage and genetic background (Glaser-Schmitt et al. 2021). Polymorphism variation also showed distinct differences in starvation resistance but not in other morphological traits such as body size (Glaser-Schmitt et al. 2021). The precise connection between this polymorphism, starvation resistance, and balancing selection is unclear, but this does show how different approaches are required to illuminate patterns of balancing selection and how these maintain genetic diversity for fitness. Future studies will also need to employ a variety of methods to uncover the identities and actions of genes contributing the genetic variation.

1.4. Knowledge gaps

Genomic approaches have been successful in identifying regions of the genome under long-term balancing selection and loci that are likely to be involved in maintaining diversity in fitness. Such studies weren't feasible until recently and are a major source of support for the role of balancing selection in evolution. However, genomic approaches do suffer from a number of issues limiting their interpretation and usefulness. The detection of balancing selection is based on identifying regions of the genome where the allele frequency spectrum has an elevated level of variation and intermediate allele frequencies, compared to that expected under neutrality or directional selection (Charlesworth 2006; Fijarcsk and Babik 2015). This method can struggle to pinpoint the loci responsible for generating balancing selection, but rather detects selection occurring in a region of the genome (Andrés et al. 2009; Bitarello et al. 2018). These patterns are also affected by factors such as drift, demography, structural variation (e.g. inversions), and linkage-disequilibrium that are not linked to balancing selection (Charlesworth et al. 1997, 2003). Although methods attempt to account for such affects, it can still be difficult to be certain that detected variation actually has an impact on fitness and is maintained by balancing selection. Even if candidates are real, the mechanism by which balancing selection occurs, e.g. sexual antagonism versus frequency dependence, cannot be determined (Fijarcsk and Babik 2015). Without phenotypic evaluation of variants we are unable to understand how a polymorphism reacts to selection and how this maintains variation. Finally, genomic approaches are poor at detecting either transient balancing selection or younger polymorphisms which have less obvious genomic signatures (Charlesworth et al. 1997; Fijarcsk and Babik 2015).

Phenotypic studies have been useful in demonstrating how variation in phenotypic traits translates to differences in fitness capable of generating balancing selection. By observing the fitness of different phenotypes we can also understand the mechanism by which balancing selection occurs. However, examples can be difficult to uncover and laborious to describe. A major issue is that studies typically describe balancing selection maintaining

genetic variation at just one locus. If balancing selection is a major factor in maintaining genetic variation across the genome, then only a few examples don't really tell us much, we required more (Bernatchez 2016). The solution is to vastly increase the number of well described examples of loci under balancing selection. However, the amount of work required to fully investigate balancing selection at the number of loci identified by genomic studies makes this impractical (Mullon et al. 2012, Ruzicka et al. 2020).

One approach that could be helpful is a combined approach using genome wide association studies (GWAS) and experiential evolution. GWAS detect associations between values of fitness traits and genetic variation. Variants consistently correlated with trait values are considered candidate loci affecting fitness in this trait. Combining GWAS and experimental evolution is effective as the GWAS identifies a set of candidate loci which can then be tracked to see how they respond to selection (Magwire et al. 2012; Martins et al. 2014; Turner and Miller 2012; Turner et al. 2013). This helps by both increasing the number of candidate loci for a trait and by providing a test that these loci respond to selection in the manner expected (Schlötterer et a. 2015). This combined approach has not yet been applied to investigate candidate loci possibly under balancing selection, but a recent GWAS by Ruzicka et al. (2019) did produce a list of candidate SA SNPs which are promising targets for balancing selection.

Experimental evolution also provides a solution to a shared problem of genomic and phenotypic approaches, that is, neither approach is good at demonstrating that a polymorphism is actively maintained by balancing selection. Genomic scans require time to build detectable genetic signatures of selection which may not represent the current situation, while newer polymorphisms will be missed. For phenotypic studies, just because a polymorphism is shown to have an effect on fitness, it does not mean that this will result in balancing selection at the population level. Experimental evolution can be combined with these approaches to potentially resolve

these issues by tracking the behaviour of a polymorphism over time (Franssen et al. 2017; Schlötterer et al. 2015).

Genomic and phenotypic approaches of balancing selection are rarely used in combination with each other, despite the complementary strengths and weakness of each approach. Therefore, the number of described polymorphisms where there is: (i) a genomic signature of balancing selection; (ii) demonstration that genetic variation directly relates to the phenotype and fitness variation; and (iii) demonstration that the polymorphism is actively maintained by balancing selection, are exceptionally rare. These three requirements need to be met for a polymorphism to be fully validated as a locus of balancing selection (Bernatchez 2016; Fijarcsk and Babik 2015). Only by identifying such loci will we be able to form a complete picture of how balancing selection contributes to maintaining genetic variation for fitness.

Additionally, future investigation should aim not just to explore balancing selection maintaining variation at a single locus but go further to add insight into the identity and biological role of such loci. Such information is required to understand how balancing selection acts across the genome to maintain genetic variation (Grieshop et al. 2021; Mérot et al. 2020; Key et al. 2014). Any new methodology that is developed should also be replicable, allowing for quicker, easier and more robust assessment of multiple candidate loci (Ruzicka et al. 2020).

1.5 Thesis overview

1.5.1 Thesis aims

To address some of these knowledge gaps, in this thesis I investigate the *D. melanogaster* gene *fruitless* as a candidate gene under balancing selection. Previous genomic work has detected an indel polymorphism in this gene which is stable at intermediate frequencies across distant populations. In this thesis I have: i) described the fitness effects of the *fruitless* polymorphism in

order to uncover how the variation at this locus impacts upon the fly's phenotype and fitness; and (ii) observed how the *fruitless* polymorphism behaves in populations and used comparative simulations to diagnose the mode of selection acting on the *fruitless* polymorphism. The result demonstrates a clear link between the *fruitless* polymorphism, fitness, and balancing selection. This fulfils the three criteria needed to confidently identify and describe fitness influencing polymorphisms which are maintained by balancing selection. I also extended the methodology of the *fruitless* analysis to detect the mode of selection acting upon candidate SA SNPs identified in the GWAS of Ruzicka et al. (2019). Doing so provides an approach to validate the balancing selection credentials of multiple SNPs at once. I hope to provide a better understanding of how balancing selection occurs, how it can be detected, the types of loci affected, and processes maintaining genetic variation.

1.5.2 Specific methodology

I will now expand on three key topics of this thesis and provide more background to them. These topics are: (i) the *fruitless* gene, where the polymorphic indel that is the subject of Chapters 2 and 3 is located; (ii) the next generation sequencing approach, pool-sequencing; (iii) the statistical method, Approximate Bayesian Computation, which was employed to diagnose the mode of selection acting at particular loci in Chapters 3 and 4. I then finish with an overview of the thesis structure.

1.5.2.1 fruitless

In Chapters 2 and 3, I focus on a polymorphism in the *fruitless* gene of *D. melanogaster* as a candidate balanced polymorphism. *fruitless* (*fru*) is a BTB-zinc-finger transcription factor gene which is highly conserved across many taxa of insects for the last 250Myr (Gailey et al. 2006). In *Drosophila melanogaster*, *fru* is an essential part of the sex-determination regulatory cascade and involved in the development of the sex-specific neural system (Ryner et al. 1996; Kimura et al. 2005). *fru* possesses 4 different promotor

regions (Anand et al. 2001) and produces several mRNA isoforms which interact with hundreds of other sites across the genome, many of which are on the X-chromosome (Neville et al. 2014; Vernes 2014). Some of these mRNA isoforms are spliced in a sex-specific manner and manage the development of the nervous system in each sex (Neville et al. 2014; Parker et al. 2014; Usui-Aoki et al. 2000). In males, the presence of the protein isoform FRU^{MC} determines the fate of neurons in the fly's nervous system (Kimura et al. 2005; Nojima et al. 2014). Females, which lack the FRU^{MC} protein experience programmed cell death of mLA neurons (Kimura et al. 2005; Sato and Yanamoto 2020). If females are forced to produce male specific isoforms they will display mating behaviours typical of males, such as attempted courtship and mating with other females (Demir and Dickson 2005). Males with mutations in the *fru* gene display altered mating behaviour including a loss of all sexual behaviours (Anand et al. 2001; Ryner et al. 1996) or alteration in courtship traits such as song (Rideout et al. 2007). Knock-out mutations in *fru* are lethal in both sexes. The strong phenotypic effects of *fru* (Anand et al. 2001; Ryner et al. 1996), the sex-specific splicing (Parker et al. 2014), and the multiple isoforms with various targets (Neville et al. 2014; Vernes 2014) makes *fru* a likely source of pleiotropic effects on various traits. This in turn makes *fru* a potential candidate for balancing selection as any genetic variation is likely to affect multiple traits in both sexes.

Previous work in our laboratory has focussed on identifying natural polymorphisms that may be targets of antagonistic selection. In connection to these efforts, previous PhD students Mark Hill and Filip Ruzicka identified a naturally occurring 43bp indel polymorphism in the regulatory region of the *fru* gene (Hill 2017, Ruzicka 2018). The polymorphism consists of the reference haplotype (S allele) and an extended haplotype carrying an insertion relative to the reference genome (L allele). Ruzicka (2018) showed that *fru* haplotype sequences from two long-time separated populations of *D. melanogaster* from the USA (Mackay et al. 2012) and Zambia (Lack et al.

2015), clustered based on their *fru* indel genotype rather than population of origin. The indel polymorphism has been stable over a long period of time, as shown by its persistence in both populations and that it exists at intermediate frequencies in them (USA: f(L)=0.475, f(S)=0.525, and Zambia: f(L)=0.511, f(S)=0.489) (Ruzicka 2018). The length of time over which this polymorphism has been preserved, the frequencies that it has been maintained at, and the pleiotropic nature of the *fru* gene, make it a promising candidate as a site where diversity is maintained through balancing selection.

1.5.2.2 Pool-sequencing

In Chapters 3 and 4 I investigate data which were generated using the sequencing technique pool-sequencing (pool-seq). Pool-seq is a genome sequencing technique that allows for highly cost-effective allele frequency data to be gathered for a group of individuals (Schlötterer et al.2014; Gautier et al. 2013). Rather than performing sequencing on samples generated from individuals, pool-seq uses DNA samples containing multiple individuals, referred to as a pool, representing a sample of the population (Futschik and Schlötterer 2010; Sham et al. 2002). Pooling thereby drastically reduces the library preparation costs of a sequencing experiment. For example if we had five populations and sampled 50 individuals from each one, we would require up to 250 DNA libraries for individual sequencing, compared to only five libraries for pool-seq. This cost effectiveness has made pool-seq an increasingly popular tool as sequencing experiments are now feasible even for those working with modest budgets (Schlötterer et al. 2014; Graves et al. 2017). Considerations when designing a pool-seq experiment include the number of individuals within each pool, ensuring that these contribute equally, and the sequencing read depth (coverage) (Schlötterer et al. 2014). Although coverage and a greater number of contributors to the pool are ideal, these need to be balanced in order to achieve an effective pool-seq experiment as both of these factors increase the volume of sequencing data required, costing more money. Due to its cost effectiveness, pool-seq is usually applied to whole-genome sequencing, but can be adapted for

targeted sequencing approaches such as exome RAD-seq strategies (Schlötterer et al. 2014).

There are a number of caveats to be aware of when designing a pool-seq experiment. The first is that since the sequenced read data comes from a mix of individuals and current sequenced read lengths are quite short (75-150bp), it is very difficult to gather any information on linkage or haplotype blocks, as there is no way of associating reads to specific individuals and hence infer linkage of SNP alleles across reads (Futschik and Schlötterer 2010; Schlötterer et al. 2014). New methods are being developed to alleviate this limitation, including large scale individual sequencing before the pool-seq experiment begins (Tilk et al. 2019) and constructing haplotype blocks from correlated allele trajectories (Michalak et al. 2019; Otte and Schlötterer 2021), but these have not yet been widely applied. Also, no information regarding the frequencies of diploid genotypes (homo-/heterozygotes) can be gathered, as SNP frequencies can only be measured for the pool as a whole (Schlötterer et al. 2014). This prevents the study of some biological processes which require information on the specific genotype of individuals, for example inbreeding using the statistic F_{IS} (Wright 1922). Despite these issues, pool-seq has been successfully applied to a variety of research areas in biology including: genome-wide association studies (Sham et al. 2002), experimental evolution (Schlötterer et al. 2015; Franssen et al. 2017), selective sweeps (Tobler et al. 2013), mapping complex traits (Ehrenreich et al. 2010), and tracking population clines (Bergland et al. 2014). In all these applications, pool-seq provides an effective way to obtain allele frequency data about a population.

1.5.2.3 Approximate Bayesian Computation

In Chapter 3 I use the statistical method Approximate Bayesian Computation to diagnose the mode of selection acting at first the *fru* locus, and then a set of 100 short intron SNPs. In Chapter 4 I employ this method again to diagnose selection at SA candidate SNPs. Approximate Bayesian

computation, or ABC, is a statistical method applied to complex datasets where traditional likelihood methods of inference are either mathematically or computationally difficult (Blum and François 2010). ABC offers a method to deal with complex biological data and test for specific effects and/or estimate statistical parameters in a manner that explicitly incorporates complexity, while at the same time avoiding the need to derive explicit likelihood functions. This is achieved by comparing simulations of a system run under a wide range of parameter values and (where applicable) different models to the real observations (Csilléry et al. 2010). Since its initial development by Pritchard et al. (1999) and Beaumont et al. (2002), ABC has been applied to answer questions in a wide range of research areas including population genetics (Sjödin et al. 2012), epidemiology (McKinley et al. 2018), and environmental modelling (Cui et al. 2018). In this thesis I apply ABC to diagnose the mode of selection acting in populations of *Drosophila melanogaster*.

ABC is mainly applied for two purposes (Csilléry et al. 2010, 2012). The first is model choice. Given a set of simulations which describes a phenomenon of interest under different models, ABC can calculate which model has the highest probability of giving rise to the real observed data. The second purpose of the ABC is parameter estimation. Simulations are built using a range of values for a set of parameters, and ABC can estimate the parameter values which would produce a simulation most closely matching the real data. To accomplish these tasks ABC uses large volumes of simulated data which describe the phenomena of interest under different conditions. ABC then compares these simulated data to the real observed data to determine the similarity between them (Csilléry et al. 2010). This comparison is performed using a number of summary statistics, which are measures taken of both the observed and simulated data, and which should ideally capture some aspect of their variation. These statistics can be formal, established statistics such as Tajima's D or Watterson's theta, or new measures specific to the experimental set up (Csilléry et al. 2010).

When comparing summary statistics ABC attempts to match the real data summary statistics to those of the simulations to find those with the closest match (Csilléry et al. 2012). For parameter estimation this is done by calculating the distance, *d*, between a simulation's value and the observed data. A tolerance rate is set so all those simulations where *d* is too large are rejected and the accepted simulations form a posterior distribution from the which the parameter values are estimated. For model choice the process is similar, but rather than estimating specific values, ABC calculates a probability of a model class generating the observed data from the proportion of simulations of that type contributing to the posterior distribution. A number of algorithms have been developed to improve this process and correct for mismatches between the summary statistics. The most commonly used are the 'rejection' method which removes one summary statistic and tests if that improves *d* or not, or regression based techniques, such as multinomial logistic regression or local linear regression, which weight the posterior distribution based on how well a summary statistic describes a parameter value or model class. The development of new algorithms is an active area of research as new methods seek to make improvements (Beaumont 2019). Increasingly these make use of machine learning algorithms such as the 'neural net' method which are better than rejection- or regression-based approaches at dealing with highly dimensional data with large numbers of summary statistics (Blum and François 2010; Csilléry et al. 2010). However, the best algorithm choice will vary depending on the data and summary statistics which is why cross-validation of approaches to estimate the errors associated with different methods is essential (Blum and François 2010). Although ABC is a powerful statistical tool it is also important to understand its limitations and not to blindly accept the results. It is therefore important to perform cross validation of all tests to have some idea of the power that ABC has and to estimate the error in its calculations (Csilléry et al. 2010, 2012). It is also important to curate the summary statistics used to ensure that all are

informative and help to differentiate the simulations performed under different models and with different parameter values.

1.5.3 Thesis structure

This thesis is composed of this introduction together with three data chapters, a discussion and three appendices. I will now give a brief description of each element that follows this introduction.

In Chapter 2, I investigate the *fruitless* (*fru*) polymorphism with the aim of characterising the fitness effects of the two alleles. This was done with six isogenic populations of *D. melanogaster*, three fixed for the S allele and three fixed for the L allele. These lines were subjected to a number of fitness assays to measure the performance of each genotype. I found that flies from lines fixed for the S allele showed improved performance relative to L flies in reproductive fitness traits in both sexes. However, L allele flies had greater survival to adulthood, and in some cases decreased likelihood of death with age. This pattern where one allele is beneficial for traits related to fitness at one stage of life and the other allele for those at another indicates antagonistic pleiotropy (AP) at the *fru* locus. I hypothesise that genetic diversity at the *fru* locus is maintained by AP generating balancing selection.

In Chapter 3, I use a suite of replicated experimental populations of flies to study the behaviour over time of the *fru* polymorphism at the population scale. Ten populations were established, five with one value of the initial frequency of the S and L alleles and five with a different initial frequency. I sampled all of these cages periodically and sequenced the fly samples using pool-sequencing. From this I obtained information about the trajectories of the *fru* allele frequencies over the course of the experiment. I use ABC to compare these trajectories to those from simulations with the *fru* locus subject to different selective conditions. From this it appears as though the *fru* locus is under some form of balancing selection, as predicted from Chapter 2.

In Chapter 4, I expand on the methodology developed in Chapter 3 to diagnose the mode of selection acting at multiple loci. The loci investigated come from a genome-wide association study (GWAS) which identified 2372 SA loci in the LHM population (Ruzicka et al. 2019). SA loci are expected to experience balancing selection and contribute to genetic variation but have traditionally been difficult to identify. Again I used ABC to compare the frequency trajectories of focal SNPs with simulation of SNPs under different selective conditions. I found that 60% of the candidate loci investigated appeared to be under some form of balancing selection, but precisely which loci experience balancing selection is difficult to predict from the GWAS results alone. I also took a closer look at five candidate loci which have been chosen to create gene edited lines using CRISPR technology.

In Chapter 5, the general discussion, I first summarise the main findings from each chapter. I then discuss my findings in relation to the field at large and what they mean for how we understand balancing selection and how genetic diversity is maintained in general. I also discuss some of the caveats of the work, suggest aspects of future study that could resolve these caveats and identify other areas of investigation that will aid in the further study of balancing selection.

This thesis also contains three appendices. Appendix A contains additional information relating to Chapter 2 about how the isogenic *fruitless* lines were created. Appendix B presents a copy of a paper and accompanying supplementary information published in Proceedings of the Royal Society B on which I was the lead author. This paper is composed mostly of work from Chapter 2. Appendix C describes additional methods and results from Chapter 3 regarding the development of summary statistics used to compare simulations by ABC.

1.6 Declaration

This PhD project was funded by a PhD studentship awarded to me by the London NERC Doctoral Training Partnership. Supervision was provided by Prof. Max Reuter and Prof. Kevin Fowler. All work presented is my own, apart from specific contributions which are described at the beginning of each chapter.

Chapter 2

2. A non-coding indel polymorphism in the *fruitless* gene of *Drosophila melanogaster* exhibits antagonistically pleiotropic fitness effects

2.1 Declaration

All work presented here is my own, except for two sections which were performed by collaborators and are included for context. Section 2.4.2, ‘Creation of allelic lines’, was conducted by Filip Ruzicka as a follow-on from his previous work identifying the *fru* polymorphism that is described in the introduction of this chapter and is described in full in Appendix A. Section 2.4.5, ‘Balancer stock genotyping’, was performed by Quentin Saintain using samples that I provided.

The work contained in this chapter was published in the Proceedings of the Royal Society B on the 12th of May 2021 (https://doi.org/10.1098/rspb.2020.2958) and this chapter is an adapted version of the published article. A copy of the article itself and its accompanying supplementary information is presented in Appendix B.

2.2 Abstract

The amount of genetic variation for fitness within populations tends to exceed that expected under mutation-selection-drift balance. Several mechanisms have been proposed to actively maintain polymorphism and account for this discrepancy, including antagonistic pleiotropy, where allelic variants have opposing effects on different components of fitness. A non-coding indel polymorphism in the *fruitless* gene of *Drosophila melanogaster* has previously been identified and here I measure survival and reproductive components of fitness in males and females of replicate lines carrying each respective allele. Expressing the *fruitless* region in a hemizygous state reveals a pattern of antagonistic pleiotropy, with one allele generating greater reproductive fitness and the other conferring greater survival to adulthood. Different fitness effects were observed in an alternative genetic background, which may reflect dominance reversal and/or epistasis. These findings link sequence-level variation at a single locus with complex effects on a range of fitness components, thus helping to explain the maintenance of genetic variation for fitness. Transcription factors, such as *fruitless*, may be prime candidates for targets of balancing selection since they interact with multiple target loci and their associated phenotypic effects.

2.3 Introduction

Genetic variation for fitness provides the raw material for selection and genetic drift to cause genetic evolution of populations (Fisher 1930). The action of both forces, however, tends to reduce genetic variation. This is particularly relevant in the case of traits that are closely linked to fitness and therefore, by definition, under strong directional selection. The classic explanation for the presence of heritable variation for fitness in populations is mutation-selection-drift balance, where standing variation is maintained at an equilibrium between the generation of new variation by recurrent mutation and its reduction through selection and drift (Muller 1950; Lewontin 1974). Yet most populations typically harbour considerable amounts of genetic variation for traits and fitness—and more than can be accounted for by mutation-selection-drift balance alone (Charlesworth 2015). This discrepancy between theoretical expectations and empirical data constitutes a central and perennial puzzle in evolutionary biology (Charlesworth and Hughes 2000; Charlesworth 2015).

One possible resolution of this paradox is that fitness variation is actively maintained by balancing selection. Initially popularised by Dobzhansky (Dobzhansky 1955), balancing selection is a force actively maintaining two or more allelic variants at a locus. The active maintenance of polymorphism requires that the selective value of an allele depends on the context in which it finds itself (Gloss and Whiteman 2016; Llaurens et al. 2017). Allelic fitness effects can depend on the genetic context within an individual, as in the case of overdominance (Johnston et al. 2013) or reciprocal sign epistasis (Ono et al. 2017) or the genetic context in the population, as with negative frequency-dependent selection (Sinervo and Lively 1996) or variable environmental conditions (fluctuating selection, (Wittmann et al. 2017)). In the case of antagonistic selection, polymorphism is maintained because the fitness effect of an allele depends on the sex of the carrier (sexual antagonism, (Kidwell et al. 1977; Bonduriansky and

Chenoweth 2009)), or on an individual's life history stage (antagonistic pleiotropy, (Williams 1957)).

Antagonistic pleiotropy (AP) occurs when mutations have a beneficial effect on one fitness component but a deleterious effect on another. Initially conceived in the 1950s (Caspari 1950; Williams 1957), AP has become a major hypothesis for the evolution of ageing, where mutations that increase fitness early in life are proposed to cause deterioration and increased mortality later in life (Williams 1957; Williams and Day 2003). AP could maintain genetic variation if, for example, one allele confers increased early-life fitness and a shorter lifespan, while the other causes a more even reproductive output over a longer life, with both strategies providing similar long-term fitness pay-offs and greater fitness than an intermediate strategy (Rose 1982; Stearns 1989). Despite some empirical evidence of pleiotropic trade-offs (Rose and Charlesworth 1981), modelling has shown that the conditions under which AP generates balancing selection and maintains polymorphism are quite restrictive (Rose 1982, 1985; Curtsinger et al. 1994; Hedrick 1999). This, combined with relatively few empirical examples of AP in nature, has led researchers to question whether AP is a major contributor to the maintenance of genetic variation for fitness (Curtsinger et al. 1994; Mérot et al. 2020).

However, recent theoretical and empirical studies have re-ignited interest in AP as a mechanism generating balancing selection. Models of metapopulation structure in fungi (Tellier et al. 2007) and viability and fertility selection in flowering plants (Brown and Kelly 2018) have demonstrated a crucial role of AP in maintaining genetic variation for fitness in wild populations. Similarly, Mérot et al. (2020) found that AP in fitness effects and the resulting variation in life-history trade-offs is most likely responsible for the maintenance of an inversion polymorphism in the seaweed fly *Coelopa frigida*. More recent theoretical models have further shown that the conditions required for AP to generate balancing selection are less stringent than

initially believed. For example, taking into account sex-specific fitness effects or even small variations in dominance between traits over time may be enough for AP to generate balancing selection under a wider range of conditions (Zajitschek and Connallon 2018). Furthermore, AP may generate excess fitness variance (relative to unconditionally deleterious mutation-selection balance) by slowing the removal of deleterious variation, rather than maintaining it *per se* (Llaurens et al. 2017; Zajitschek and Connallon 2018). Together these developments suggest that the proportion of AP genetic variation (and possibly balanced variation) has been historically under-estimated (Charlesworth 2015), underscoring the need for further experiments that link sequence-level polymorphism with measurements of fitness components at different life stages, ideally in both sexes.

Here, I investigate fitness effects associated with a polymorphism in a non-coding region of the *fruitless* gene (*fru*) of *Drosophila melanogaster*. The *fru* gene is a key component of the sex-determination cascade and is responsible for sex-specific nervous system development and courtship behaviour (Ryner et al. 1996; Kimura et al. 2005; Neville et al. 2014). In line with its crucial functions, *fru*'s protein coding sequence is conserved across insect taxa (Gailey et al. 2006). Contrasting with the evolutionary constraint that is evident at the phylogenetic level, *fru* also exhibits evidence of positive selection (Parker et al. 2014).

Previous work in this laboratory by a former PhD student (Filip Ruzicka) identified a polymorphism within the 5' non-coding region of the *fru* gene (Ruzicka 2018). This was achieved by investigating signatures of balancing selection in population genomic data from two collections of wild flies from Raleigh, US (MacKay et al. 2012) and Zambia (Lack et al. 2015), using metrics of genetic diversity (nucleotide diversity, Tajima's D) and linkage disequilibrium, which can indicate if a region is under balancing selection (Fijarczyk and Babik 2015). Sanger sequencing of a 400bp region of LH_M (a laboratory-adapted population (Rice et al. 2005)) revealed a 43bp

indel polymorphism, (henceforth referred to as the *fru* indel) that is in perfect LD with seven SNPs in the flanking sequence (Figure 2.1A). A haplotype network constructed from these SNPs showed that haplotypes do not cluster by population but fall into divergent allelic classes that occur at intermediate frequencies in both populations (Figure 2.1C). Given the large evolutionary distances between the Raleigh and Zambia populations, this is suggestive evidence that the *fru* indel (and/or alleles linked to it) is under some form of antagonistic and/or balancing selection.

To investigate why this locus is consistently polymorphic, I designed and performed fitness experiments to assess the consequences of bearing each respective allele for multiple fitness components in both sexes. I found that one allele confers higher reproductive fitness in both sexes, while the alternative allele results in greater larval survival and, in some cases, greater adult longevity. These effects further depend on the genetic background in which the alleles are expressed, suggesting that dominance reversal and/or epistasis may also contribute to the maintenance of this polymorphism. This study adds to the growing body of evidence for a reassessment of the role played by antagonistic pleiotropy, and possibly balancing selection, in maintaining individual allele polymorphisms and genetic variation for fitness.

2.4 Methods

2.4.1 Fly culture and husbandry

Unless otherwise stated, flies were maintained on corn-agar-molasses medium with a powdering of live yeast in either vials (8ml of media) or bottles (50ml) in 25°C constant temperature rooms at 50% humidity on a 12:12hr light-dark cycle. When required, flies were collected as virgins, every 0-6 hours post-eclosion until sufficient numbers were obtained. Flies were anaesthetised using a CO_2 pad for short periods of time and manipulated using a fly aspirator or paint brush.

2.4.2 Creation of allelic lines

Isogenic allelic lines, which were fixed for either the S or L allele but otherwise isogenic for a Canton-S background were created in order to study the fitness effects of the *fru* polymorphism. Details for this process are described in Appendix A.

2.4.3 Generating focal flies

Fitness assays were performed on "focal" flies generated by crossing individuals from the allelic lines to flies from the *Df(3R)fru*$^{4\text{-}40}$*/TM6B* stock. The resulting individuals carried the *fru* allele (L or S) of a line complemented either by the *Df(3R)fru*$^{4\text{-}40}$ deficiency (D) or by the TM6B balancer chromosome (B). Since the deleted region of the *Df(3R)fru*$^{4\text{-}40}$ chromosome extends over the *fru* locus, flies which inherit this chromosome (D) are hemizygous for whichever *fru* allele they inherit. The *fru* alleles can therefore be studied in isolation in D flies. The B chromosome (TM6B) was found to carry the S allele. The contrast of allelic fitness effects between flies complemented with the D deficiency or the B chromosome thus allows us to gain information on dominance effects of the *fru* alleles and epistatic interactions with the genetic background. The cross to generate focal flies also ensures that line-specific recessive deleterious alleles are masked by complementing with both B and D chromosomes, so as to minimally affect

fitness measurements associated with the *fru* alleles. Before crossing, flies were maintained for multiple (>10) generations in bottles containing molasses media, at a population size of 200-300 flies per bottle and 3 bottles per line.

For each line (S1–3 and L1–3), crosses were performed by setting up replicate vials containing 10 virgin allelic line females and 10 *Df(3R)fru*$^{4-40}$*/TM6B* males. These vials were left overnight for the flies to mate. To limit larval densities, I twice transferred flies to fresh vials for 4-hour egg lays (~10am–2pm and ~2–6pm). To establish focal flies carrying the *fru* allele paired with either the D complement (wildtype pupal phenotype) or the B complement (*Tb* pupal phenotype), emerging pupae were sorted into separate vials based on their phenotype. Twelve total line sets were thus established, i.e. lines S1–3 and L1–3 in D or B background, referred as S/D, S/B, etc. when referring collectively to all 3 lines carrying a particular allele.

2.4.4 Fitness assays

2.4.4.1 Reproductive success

Focal females were mated to males from their own vial before being placed as triplets at 3 days old into vials containing 1% agar and fed by a capillary tube through the stopper containing a 4:1 yeast to sugar solution (6.5g yeast extract and 1.625g sugar per 100ml) at 25°C and 80% humidity, with new food capillaries supplied daily. Triplets were maintained until the focal females were 4–5 days old, since females are initially reluctant to lay in this novel environment and need time to grow accustomed to it. Triplets were then transferred to new agar vials (this time 0.8% agar was used since a lower agar % produced clearer photos) at ~4pm and allowed to lay eggs for 18 hours. Vials were photographed using webcamSeriesCapture (github.com/groakat/webcamSeriesCapture) software and a Logitech HD Pro webcam C920. I used the machine learning programme *QuantiFly* (github.com/dwaithe/quantifly) (Waithe et al. 2015) to count the eggs in each picture. Vials where a female died or where bubbles, debris, or other

contaminants caused counting problems were removed from further analysis. Fitness was assayed in 3 experimental blocks. In total, 863 successful female fecundity trials were performed.

Focal males were reared on standard food in vials of 30 mixed sex flies until 4–5 days old. To assay male mating success, focal males were paired with a competitor male from the *Df(3R)fru$^{4\text{-}40}$/TM6B* stock. Pairs of males were held in vials overnight. The next morning a virgin *Df(3R)fru$^{4\text{-}40}$/TM6B* female was added to the vial without CO_2 anaesthesia and the two males competed for mating. The males were allowed to compete for 90mins, thereby maximising the likelihood of a single mating while keeping the rate of double matings negligible. The males were then removed and the female left to lay eggs over a period of several days. Once the larvae pupated, paternity was scored using the pupal phenotype. If all pupae displayed the *Tb* phenotype then paternity was assigned to the competitor (*Df(3R)fru$^{4\text{-}40}$/TM6B*) male. If pupae were a mixture of wildtype and *Tb*, paternity was assigned to the focal male. Only vials with >10 pupae were included in further analysis, to ensure that the probability of not observing any wildtype pupae among the offspring of a wildtype male would be minimal (0.5^{10} = 0.001) and paternity could be reliably scored. In total, I obtained data on mating success for 1149 males across 3 experimental blocks.

2.4.4.2 Larval survival, sex ratio and development time

Fifty virgin females from the *fru* allelic lines and fifty males from the *Df(3R)fru$^{4\text{-}40}$/TM6B* line were placed together into egg-laying chambers (~2.5cm diameter, 5cm height) to mate and lay eggs. The floor of these chambers was composed of a grape juice/agar mixture (172ml concentrated grape juice per litre) with a small quantity of yeast as a protein source. After 48 hours, once they had acclimatised to the conditions, the flies were transferred to an identical chamber with the same food source and left for a further 24–30 hours to lay the eggs which would become the “focal” larvae assessed in this assay. Newly hatched, 1st instar larvae were picked and

placed in groups of 50 into vials containing standard medium and left to develop. Newly formed pupae were removed from the vial and placed into new vials depending on their phenotype (*Tb* or wildtype). For each vial and line, the number of eclosing flies of each sex, the proportion of surviving larvae, and the sex ratio (once all flies eclosed) was recorded. Development time was recorded as the number of days from when larvae were placed in the vial until eclosion as an adult. Complete data on larval survival, sex ratio and development time was collected for 2052 flies (1049 females and 1003 males) from 180 vials.

2.4.4.3 Lifespan

Due to the larger number of flies required for this assay compared to previous assays, focal flies were generated using a slightly different method. Groups of 100 *fru* allelic line females and 100 *Df(3R)fru*$^{4\text{-}40}$*/TM6B* line males were placed together in an enclosure containing a petri dish filled with corn-agar-molasses medium and left to lay eggs overnight. The next day, small sections of the media, each containing a similar number of eggs, were cut out and placed into individual vials. The eggs were then left to hatch and the larvae to develop. As pupae emerged the flies were separated into vials depending on the pupal phenotype (*Tb* or wildtype). The vials were checked daily until sufficient flies for the experiment eclosed on the same day, which occurred 10 days after eggs were laid. All flies used in the assay were virgins and varied in age by no more than 24 hours. Newly eclosed flies were anaesthetised with CO_2, separated by sex, and placed in vials in groups of 10. Every other day (Monday, Wednesday, Friday), flies were transferred to a new vial without anaesthesia. The number of dead flies at each transfer was recorded and dead flies removed. If a fly escaped this was recorded and included in the analysis by censoring. This process was continued until all flies had died. Complete lifespan data was collected for 1659 flies, with partial data obtained for another 257 flies.

2.5.5 Balancer stock genotyping

In order to asses any sex- or trait-specific dominance of the *fru* alleles I needed to know the allele carried by the TM6B balancer chromosome. DNA was extracted from several *Df(3R)fru*$^{4\text{-}40}$*/TM6B* flies and the indel region was then PCR-amplified. The size of the PCR product was checked on an agarose gel (using control reaction with L- and S-bearing DNA templates from our isogenic lines as controls) and Sanger-sequenced to confirm allelic identity.

2.4.6 Statistical analyses

All statistical analyses were performed in *RStudio* (R Core Team 2019). Mixed effects models were fitted using the package *lme4* (Bates et al. 2015). All mixed effects models included the flies' line ID (S1–3 or L1–3) as a random variable. If the assay was carried out in multiple blocks, this was also included as a random effect. P-values for each model term were calculated using parametric bootstrapping (package *pbkrtest* (Halekoh and Højsgaard, 2014)) based on 1000 simulations.

Egg count output from the *QuantiFly* programme was square root transformed (to achieve better model fitting) and analysed using a linear mixed effects model (LMM) with Gaussian error. The model included the *fru* allele (L or S), chromosomal complement (B or D) and their interaction as fixed effect parameters.

Male mating success was recorded by scoring paternity (focal vs. competitor male) as a binary response variable. A GLMM (generalised linear mixed effects model) with logit link function and binomial error structure was then fitted for this variable, containing the male's *fru* allele, its chromosomal complement, and the interaction between the two, as fixed effects. This assay also included experimental block as a random effect in the model.

Larval survival was measured as the number of adult flies emerging from each vial. An LMM with Gaussian error was applied to the log-transformed number of surviving offspring as a response variable. This produced a better fit according to log-likelihood and AIC than using a GLMM with a Poisson error distribution. The offspring's *fru* allele and chromosomal complement were included in the model as fixed effects. An additional random variable was added to account for the identity of the vial housing each fly before separation at the pupal stage.

Sex ratio was calculated as the number of males divided by the total number of flies which emerged from each vial and square-root transformed. A Gaussian LMM was applied to the sex ratio values which included *fru* allele and chromosomal complement as fixed effects and an additional random variable to account for differences between individual vials.

Development time was analysed using a Gaussian LMM including *fru* allele, chromosomal complement, sex and their interactions as fixed effects and larval vial and fly line as random effects. Development time was log-transformed to improve the model fit.

Lifespan data was analysed using Cox proportional hazard models (CPH) from the R package *survival* (Therneau 2015). A model was constructed including *fru* allele, sex and chromosomal complement as explanatory variables. Significance of model terms was assessed with sequential likelihood ratio tests. Additional models were run with single explanatory variables on either the entire or stratified datasets to estimate hazard ratios for significant model terms. Kaplan-Meier survival curves were fitted using functions from the *survminer* package (Kassambara et al. 2019).

2.5 Results

2.5.1 Reproductive success

There was no effect of the *fru* allele alone on the number of eggs laid (χ_1^2=2.62, p=0.189; Figure 2.2A). However, there was an effect on fecundity due to the chromosomal complement, with D females laying 7.3% more eggs than B females (χ_1^2=4.31, p=0.041; Figure 2.2A). Furthermore, there was a significant allele-by-complement interaction, whereby S/D flies laid more eggs (21.6% excess) than all other genotypes (χ_1^2=4.29, p=0.031; Figure 2.2A).

There was no effect of the *fru* allele on male mating success (χ_1^2=0.49, p=0.562; Figure 2.2B). The success rate of B males was 32.5% higher than that of D males (χ_1^2=17.38, p=0.001; Figure 2.2B). There was a difference between the alleles when in a hemizygous state (D complement) with S/D males achieving 35.8% more matings than L/D males, though the allele-by-complement interaction was nearly statistically significant (χ_1^2=3.52, p=0.058).

2.5.2 Larval survival and sex ratio

A greater number of L allele larvae survived to adulthood compared to S allele larvae (a 51.2% survival benefit of the L allele; χ_1^2=7.64, p=0.016; Figure 2.3) and more larvae inheriting the D chromosome survived to adulthood than those inheriting the B chromosome (22.56% more D than B larvae survived; χ_1^2=17.95, p<0.001; Figure 2.3). There was no evidence for an interaction between *fru* allele and chromosomal complement (χ_1^2=1.25, p=0.275; Figure 2.3). There were also no significant effects on the sex-ratio of emerging adult flies due to either *fru* allele (χ_1^2=0.054, p=0.809), chromosomal complement (χ_1^2=2.14, p=0.158) or their interaction (χ_1^2=2.89, p=0.097; Figure 2.4).

2.5.3 Development time

Females developed 2.1% faster than males across all genotypes (χ_1^2=98.69, p=0.001, Figure 2.5) and the B chromosome led to faster development than the D chromosome by 2.5% (χ_1^2=9.21, p=0.003). Yet, the *fru* allele had no significant effect on development time (χ_1^2=0.36, p=0.655), nor was there support for two-way interactions between any of the variables (allele-by-sex: χ_1^2=0.91, p=0.357; allele-by-chromosome: χ_1^2=0.038, p=0.848; chromosome-by-sex: χ_1^2=2.52, p=0.106) nor between all three variables (χ_1^2=0.012, p=0.921) (Figure 2.5).

2.5.4 Lifespan

A global analysis across the entire dataset did not reveal a significant effect of allele (p=0.71; Figure 2.6). However, a significant effect of complement was not found (p<0.001), with greater lifespan (smaller hazard) in flies with the D than the B complement ($HR_{D/B}$=0.72), and sex (p<0.001), with greater lifespan in males ($HR_{M/F}$=0.82). The latter effect is probably largely driven by a significant complement-by-sex interaction (p<0.001), where the direction of the sex-difference in survival is reversed between the D complement ($HR_{M/F}$=1.27) and the B complement, with a large drop in survival of B females ($HR_{M/F}$=0.50, Figure 2.6). In addition, significant pairwise interactions between allele and complement were found (p=0.001; D complement: $HR_{S/L}$=0.84; B complement: $HR_{S/L}$=1.14) and between allele and sex (p=0.028; females: $HR_{S/L}$=1.04; males: $HR_{S/L}$=0.93). The three-way interaction was not significant (p=0.25). For full survival results see Table 2.1.

2.6 Discussion

In this chapter, I describe the fitness consequences of variation at an indel polymorphism in the fruitless gene, previously identified by Ruzicka (2018), by measuring the performance of allelic lines for a number of relevant fitness components, in both sexes. The data provide evidence for complex allelic fitness effects (see Table 2.2 for a summary), with variation in the impact of the *fru* alleles between fitness components, sexes and chromosomal complements.

For cases where the *fru* allele was present in a hemizygous state, i.e. paired with the D chromosome, the effects are compatible with AP, in which alleles affect fitness in different and opposing ways (Table 2.2). Thus, flies inheriting the S allele outperformed L flies in assays of male and female adult reproductive fitness, with S females laying more eggs than L females and S males tending to have greater competitive mating success than L males. Conversely, flies inheriting the L allele had greater larval survival than those with the S allele in both sexes. These contrasting effects on reproductive fitness and survival suggest that allelic variants at the *fru* locus act antagonistically, contributing to a major life history trade-off.

In addition to AP effects, I also find evidence for interactions between the focal *fru* alleles and their chromosomal complement, which is either a wildtype chromosome carrying the deficiency $Df(3R)fru^{4\text{-}40}$ (D) or a balancer chromosome *TM6B* (B). Because the latter carries an S allele, such that L/B flies are L/S heterozygotes while S/B flies are S/S homozygotes, the comparison between the genotypes in the two complements allows us to make some inferences about dominance. Estimates of phenotypic means from the data suggest dominance for two traits, male mating success and larval survival. For male mating success, S/B (S/S) and L/B (L/S) males perform equally well while S/– males have greater mating success than L/– males (Figure 2.2B, significant allele-by-complement interaction), possibly

suggesting dominance of the S allele. For larval survival, in contrast, the difference in eclosion rate between S/S and S/L individuals is similar to the difference between S/– and L/– individuals (Figure 2.3; significant allelic effect but no allele-by-complement interaction), suggesting that the L allele is dominant for this phenotype. These potential findings of trait-specific dominance raise the intriguing possibility of adaptive dominance reversal, where the beneficial allele is dominant for both traits.

Yet there is also evidence for more complex genetic interactions. There was no difference between the effect of the two alleles on adult mortality when paired with the D chromosome, but in females L flies had lower adult mortality than S flies when paired with the B chromosome. This pattern is indicative of epistatic interactions between the focal polymorphism and the genetic background (as well as the sex-determining pathway). It is not surprising that such interactions should be apparent in these data, given the large number of sequence differences that will be present between the B and D chromosomes. What is less clear is to what degree these effects are biologically meaningful, given the presumably unnaturally high deleterious mutation load on the balancer chromosome. Nevertheless, the fact that epistatic allelic differences for particular fitness components arise in the presence of both complements makes it plausible that similar, albeit potentially weaker, effects would occur in interactions of *fru* alleles with naturally occurring polymorphisms elsewhere in the genome.

Life-history traits, such as adult fecundity and survival probability (Rose 1982, 1985) that I measured here, are often thought to be associated with genetic trade-offs (Stearns 1989). In such cases, an increase in performance in one fitness component leads to concurrent decreases in performance in another, for example due to resource allocation. Within this framework, AP is likely to occur when mutations affect the allocation that underlies the trade-off. AP effects can sometimes maintain genetic polymorphism in general models (Rose 1982, 1985), models replicating the

properties of specific natural systems (Tellier et al. 2007; Brown and Kelly, 2018) and in empirical observations (Mérot et al. 2020). Similarly, the antagonistic fitness relationship uncovered between the two *fru* alleles may maintain genetic variation at the *fru* locus.

Supporting this interpretation, the findings presented here contradict some of the assumptions used when arguing against a plausible role of AP in maintaining polymorphism through balancing selection (Curtsinger et al. 1994; Hedrick 1999). For example, classic theory predicts that in order for AP to maintain polymorphism, fitness effects need to be large and similar across fitness components, leading to doubts about the ability for AP as a source of balancing selection based on the assumption that fitness effects are small (≤1%) in most cases (Curtsinger et al. 1994; Charlesworth and Hughes 2000). Interestingly, however, the fitness differences observed is considerable. In D flies, where AP is evident, S females lay 25.1% more eggs than L females (29.67 versus 23.57) and S males achieve a third more matings than L males (40% versus 30%), while L flies of both sexes survive to adulthood with a probability that is 46.5% greater than that of S flies (14.62% versus 9.98%). The efficacy of AP-selection would also be weakened if fitness effects were limited to one sex (Curtsinger et al. 1994; Hedrick 1999). But this again is not the case here: we observe similar effects in both sexes for both reproductive fitness and egg-to-adult survival, although no reversal of fitness effects is found between the sexes (sexual antagonism), which could have further facilitated maintenance of polymorphism in conjunction with AP (Zajitschek and Connallon 2018). Another property that aids the maintenance of polymorphism via AP is dominance reversal, where the beneficial effect of each allele on a given fitness component is dominant (Hedrick 1999). Interestingly, this data provides some evidence for such a pattern, with the S allele exhibiting a dominant beneficial effect on male mating success, while the L allele exhibits a dominant beneficial effect on larval survival (see Figures 2.2B and 2.3 and discussion above). However, conclusions of trait specific dominance should

be treated cautiously since these cannot be tested fully without the the full complement of possible genotypes. The aggregate heterozygote advantage produced by these two effects will generate balancing selection that helps stabilise the polymorphism at *fru*. In addition, genetic variation could be further stabilised by epistatic interactions (Llaurens et al. 2017) such as those observed in fly survival (Figure 2.6) and discussed above. Theoretical models don't often consider epistatic effects in regards to AP, but models have shown that epistasis can help maintain polymorphism at sexually antagonistic loci (Arnqvist et al. 2014) and similar processes could, in principle, affect AP loci.

Beyond evolutionary dynamics, these results raise the question of how genetic variation at the *fru* locus generates phenotypic effects across the different fitness components measured. The FRU protein is a BTB-zinc-finger transcription factor and is produced in multiple isoforms, some of which are sex-limited (Ryner et al. 1996; Anand et al. 2001; Neville et al. 2014). The sequence differences between the L and S alleles are upstream of the coding regions, close to the sex-specific promotor P1. Accordingly, the differences observed here between the alleles must arise due to differences in expression levels rather than coding changes, and potentially due to the relative concentrations of different sex-limited and shared isoforms. Both the absolute and relative concentrations of different isoforms could potentially have important consequences on organismal function and phenotypes, given *fru*'s role as a top-level transcription factor. The number of its targets (between 217–291 depending on the particular isoform, (Vernes 2014)) would be expected to generate considerable trickle-down effects through the regulatory cascade. Even slight initial differences in *fru* expression between L and S alleles could potentially result in major, and pleiotropic, effects on a range of phenotypes. For example, mutations in *fru* can result in drastic changes in male mating behaviour and brain development (Kimura et al. 2005; Neville et al. 2014; Nojima et al. 2014). The large number of target sites also provides a potential mechanism for the epistatic interactions

observed, depending on the interplay between the abundance of the different FRU isoforms, the specific sites they bind to and the regulation that results from that binding. It is difficult to make inferences about these regulatory effects. But investigation of the sites which interact with fruitless is ongoing (Vernes 2014) and together with a more detailed knowledge of how the target loci are involved in behavioural and morphological traits, this will shed light on the mechanism(s) that link *fru* to downstream traits.

In addition to the effects of allelic variants, complements and their interaction, a significant amount of fitness variation between individual lines carrying the same genotype is observed. The method of introgression used to create the allelic lines involved naturally occurring, stochastically placed break points. As a consequence, introgressing a specific allelic variant into the region of interest will also introduce some flanking sequence of unknown size. Variation in the extent of that flanking sequence can generate differences in phenotype between lines carrying a given genotype in the target region. In principle, variation in flanking sequence could also produce systematic differences between S and L lines. In this case, however, the causative variation would require high LD with the S and L alleles.

Notwithstanding these caveats, this study provides a rare manipulative experimental test of the hypothesis that AP maintains polymorphic variation at an individual candidate gene. These results provide evidence for allelic variants at the *fru* locus generating AP between fitness components where one allele (L) enhances survival and the other allele (S) enhances reproduction. Since the *fru* polymorphism influences multiple fitness components, and each allele is beneficial in some instances and deleterious in others, the data supports the idea that the *fru* polymorphism is maintained through large antagonistic effects on fitness components, in conjunction with dominance reversal. The results complement recent findings in other systems (Mérot et al. 2020), indicating that AP is a plausible mechanism for maintaining genetic variation for fitness.

2.7 Figures and Tables

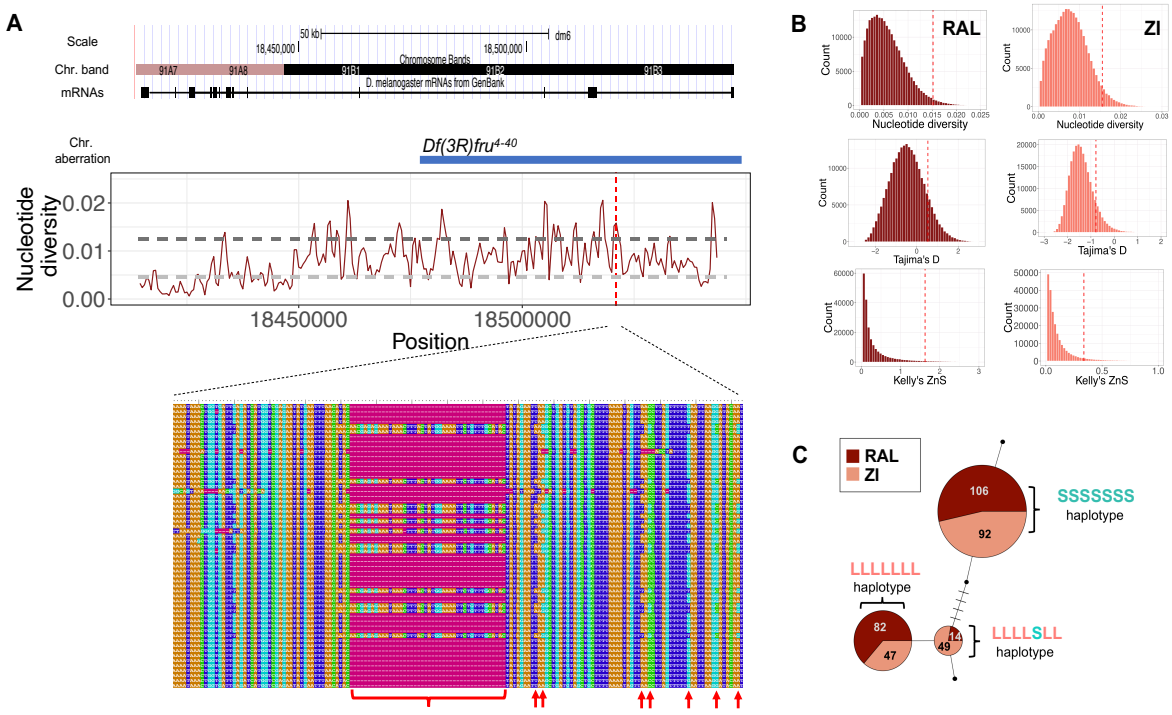


Figure 2.1. Population genetic signatures of polymorphism in the *fru* gene. **A)** Map of the *fru* gene, including breakpoints of chromosome bands, gene model, approximate span of the *Df(3R)fru*$^{4\text{-}40}$ deletion, nucleotide diversity (in RAL) in 1000bp windows (grey horizontal lines = median genome-wide nucleotide diversity; dark grey horizontal lines = 95% quantile of genome-wide nucleotide diversity) and position of the *fru* indel (vertical red dashed line). Alignments of a subset of the ~400bp region spanning the *fru* indel (brackets) obtained through Sanger sequencing of LH_M-derived chromosomes are also shown, with closely linked SNPs (used to construct the haplotype network shown in C.) shown as red arrows. **B)** Histograms of nucleotide diversity, Tajima's D and Kelly's ZnS for all 1000bp windows across the genome in RAL and ZI populations, with the vertical red dashed line representing the 1000bp window encompassing the *fru* indel. **C)** Haplotype network constructed from SNPs closely linked to the *fru* indel (red arrows in A.) in RAL and ZI populations.

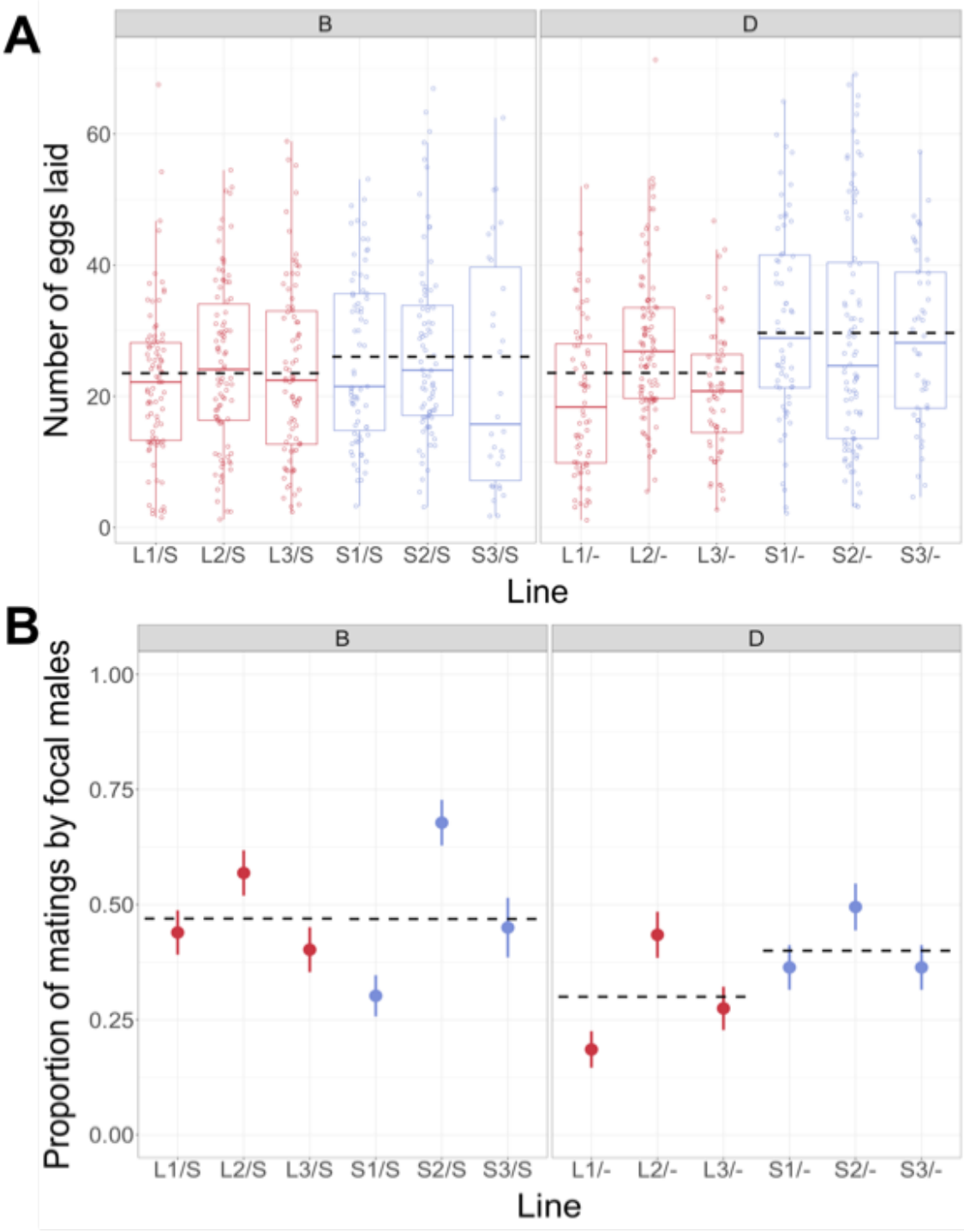

A
B
D
Number of eggs laid
0
20
40
60
L1/S
L2/S
L3/S
S1/S
S2/S
S3/S
L1/-
L2/-
L3/-
S1/-
S2/-
S3/-
Line
B
Proportion of matings by focal males
0.00
0.25
0.50
0.75
1.00
Line

Figure 2.2. The fru polymorphism and reproductive fitness traits. **A)** Number of eggs laid by triplets of focal females from each line (L1-3 and S1-3) and chromosomal complement (B and D) over an 18-hour period. Allelic means represented by dashed lines (L/B: 23.57±0.79; S/B: 26.03±1.06; L/D: 23.57±0.78; S/D: 29.67±1.13). **B)** Proportion of matings (±standard error) obtained by focal males for each line (L1-3 and S1-3) and chromosomal complement (B and D). Allelic means represented by dashed lines (L/B: 0.47±0.028; S/B: 0.468±0.031; L/D: 0.299±0.027; S/D: 0.407±0.029). Individual data points are not shown in panel B, as the response is binary (taking only values of 0 and 1).

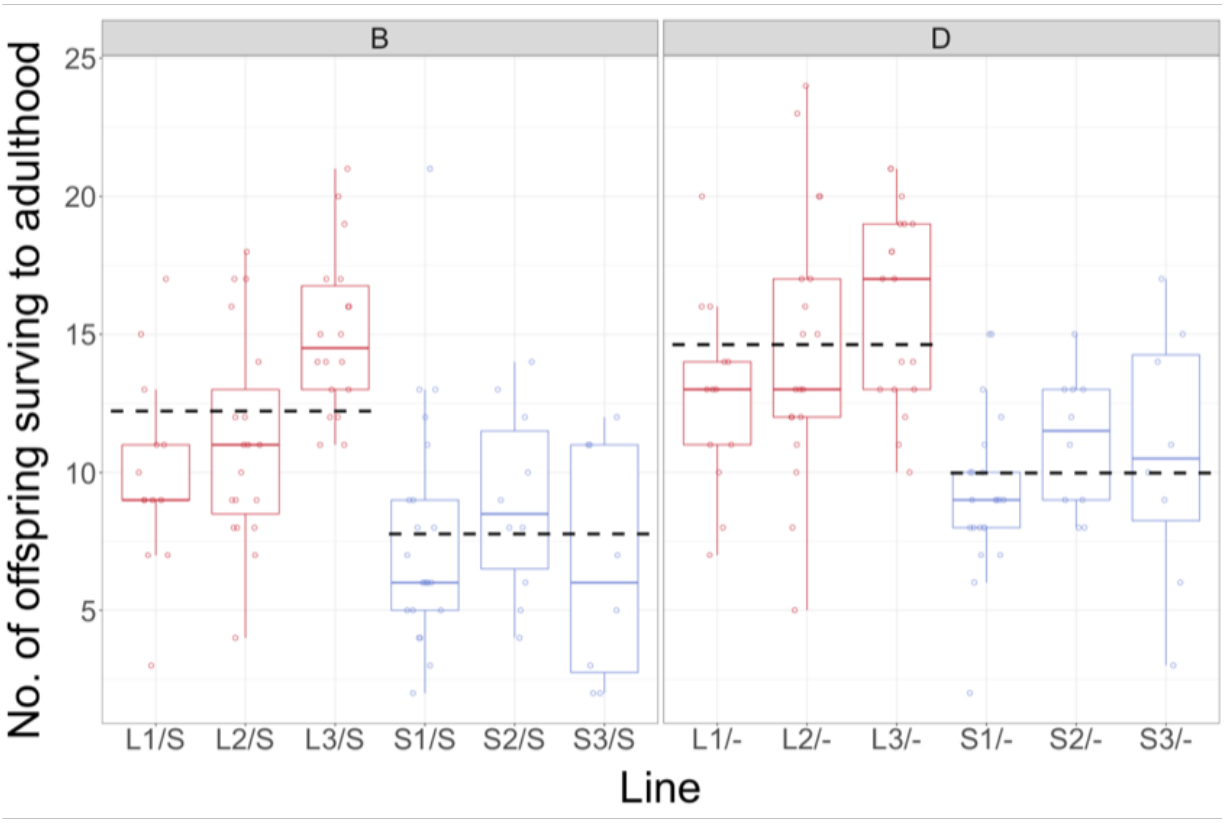


Figure 2.3. Number of offspring surviving from egg to adulthood for each line (L1-3 and S1-3) and chromosomal complement (B and D). Allelic means represented by dashed lines (L/B: 12.22±0.57; S/B: 7.78±0.64; L/D: 14.62±0.6; S/D: 9.98±0.51).

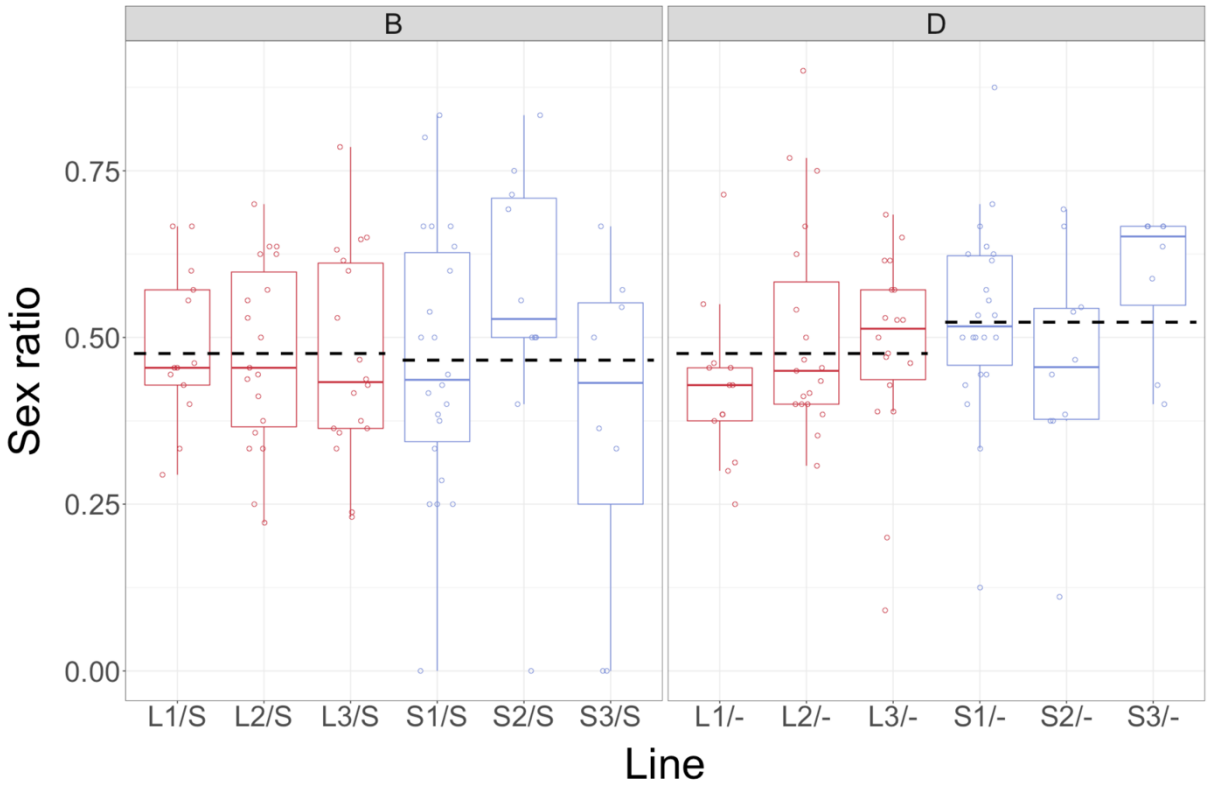
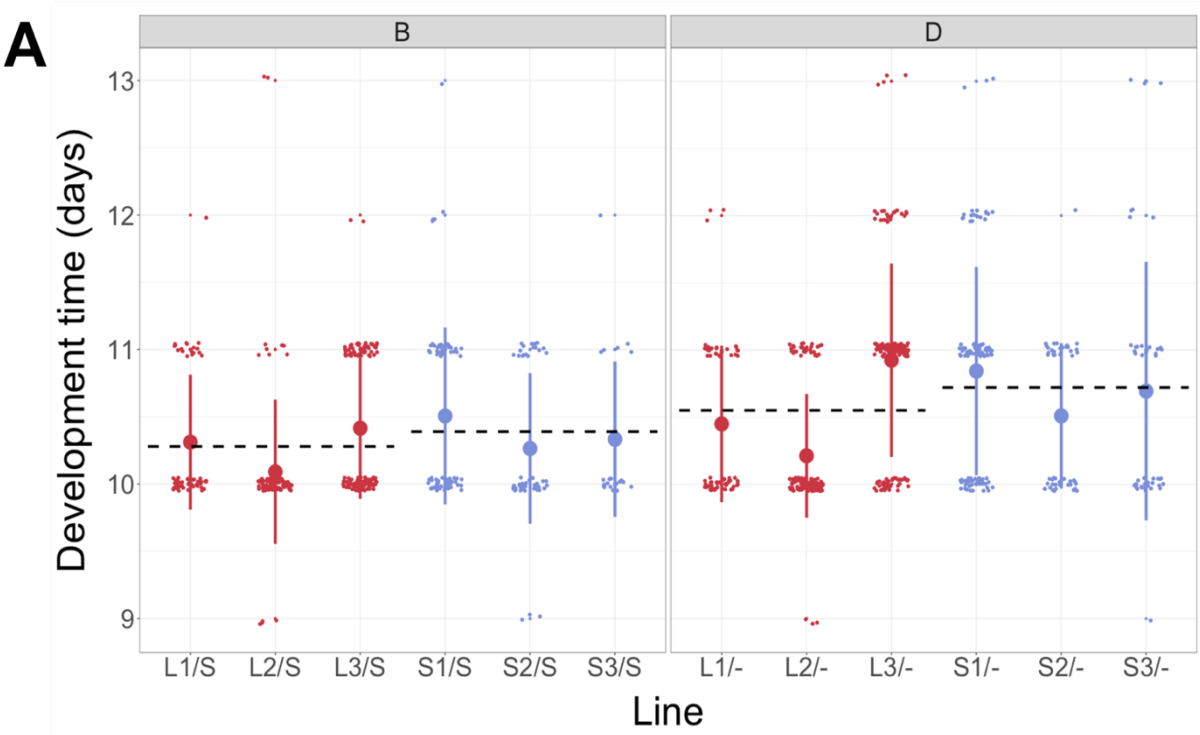


Figure 2.4. Sex ratio among surviving offspring presented for each line (L1-3 and S1-3) and chromosomal complement (B and D). Allelic means represented by dashed lines (L/B: 0.476±0.019; S/B: 0.466±0.035; L/D: 0.477±0.021; S/D: 0.0523±0.024). Sex ratio is defined as the proportion of males among offspring at eclosion.


A
B
D
Development time (days)
13
12
11
10
9
L1/S
L2/S
L3/S
S1/S
S2/S
S3/S
L1/-
L2/-
L3/-
S1/-
S2/-
S3/-
Line

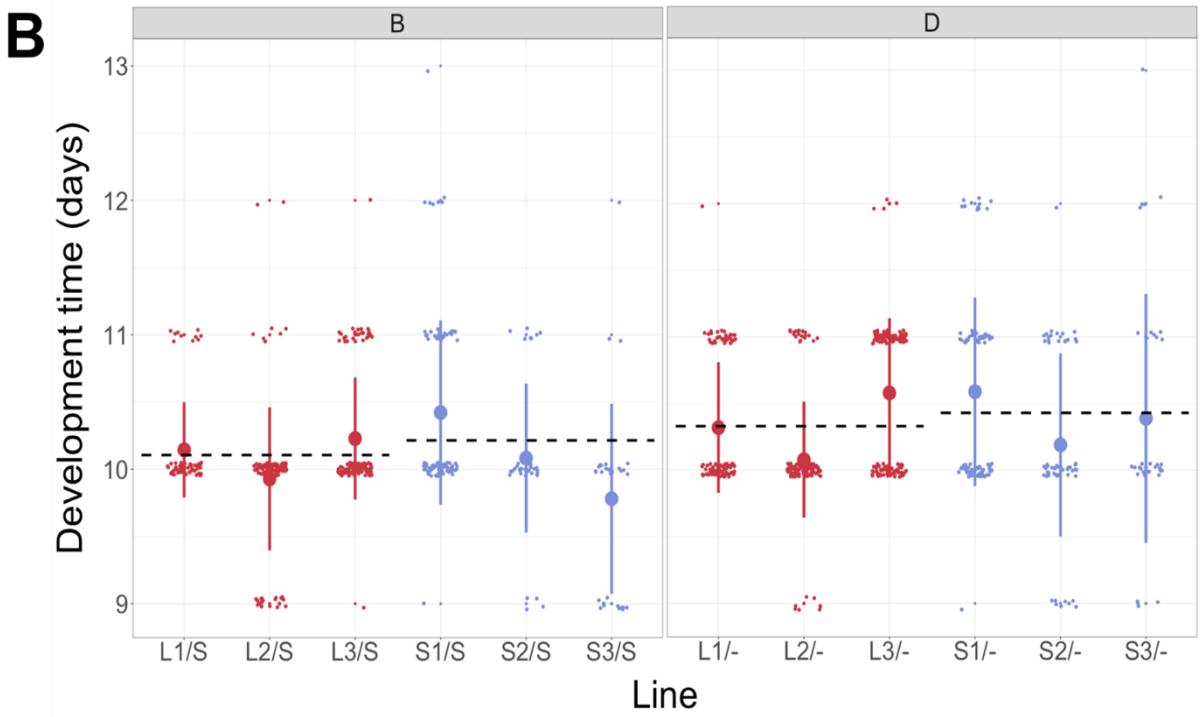

B
B
D
Development time (days)
13
12
11
10
9
L1/S
L2/S
L3/S
S1/S
S2/S
S3/S
L1/-
L2/-
L3/-
S1/-
S2/-
S3/-
Line

Figure 2.5. Development time (days ±standard error) of *fru* allelic lines (L1-3 and S1-3), for each chromosomal complement (B and D). Allelic means represented by dashed lines. Since sex was the most important factor in determining development time, this data is presented for the sexes separately: **A)** male flies (L/B: 10.28±0.03; S/B: 10.4±0.05; L/D: 10.55±0.036; S/D: 10.72±0.054), and **B)** female flies (L/B: 10.1±0.027; S/B: 10.22±0.056; L/D: 10.33±0.028; S/D: 10.42±0.056).

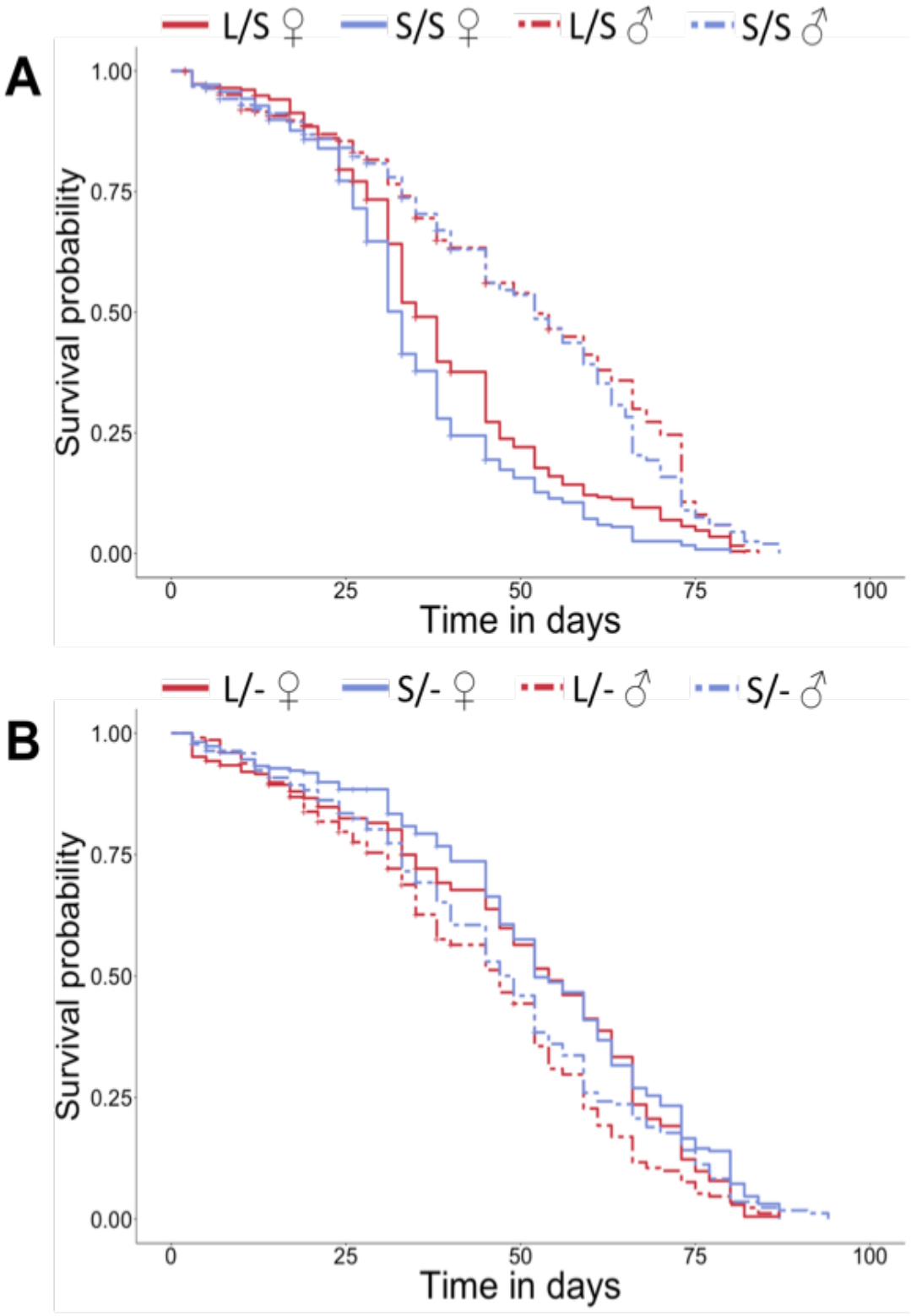

A
L/S ♀
S/S ♀
L/S ♂
S/S ♂
Survival probability
1.00
0.75
0.50
0.25
0.00
0
25
50
75
100
Time in days
B
L/- ♀
S/- ♀
L/- ♂
S/- ♂
Survival probability
1.00
0.75
0.50
0.25
0.00
0
25
50
75
100
Time in days

Figure 2.6. Kaplan-Meier survival curves of flies carrying the B complement **(A)** and D complement **(B)**. Line colour designates *fru* genotype (red = L allele, and blue = S) and line type indicates sex (solid line = females and dashed line = males). For example, the blue dashed line represents S allele males.

Model	Term (comparison)	HR	95%-CI	HR p-value	χ_1^2	p-value
All flies	*fru* allele (S:L)	1.318	1.126-1.544	<0.001	0.139	0.71
	Chro. (D:B)	0.519	0.44-0.612	<0.001	43.79	<0.001
	Sex (M:F)	0.531	0.449-0.627	<0.001	31.886	<0.001
	Allele x chro. (S/D:L/F)	0.693	0.57-0.841	<0.001	10.411	0.0013
	Allele x sex (S/D:L/B)	0.821	0.676-0.997	0.046	4.856	0.0276
	Chro. x sex (D/M:B/F)	2.624	2.154-3.198	<0.001	90.752	<0.0001
	Allele x chro. x sex (S/D/M:L/B/F)	1.258	0.852-1.856	0.249	1.331	0.249
B only	*fru* allele (S:L)	1.386	1.16-1.655	<0.001	3.848	0.049
	Sex (M:F)	0.572	0.472-0.692	<0.001	105.65	<0.001
	Allele x sex (S/D:L/B)	0.731	0.561-0.953	0.02	5.368	0.021
D only	*fru* allele (S:L)	0.87	0.715-1.059	0.164	5.317	0.021
	Sex (M:F)	1.32	1.081-1.614	0.0066	10.705	0.001
	Allele x sex (S/D:L/B)	0.927	0.696-1.234	0.604	0.269	0.604
Females only	*fru* allele (S:L)	1.381	1.157-1.65	<0.001	2.334	0.127
	Chro. (D:B)	0.542	0.449-0.655	<0.001	14.879	<0.001
	Allele x chro. (S/D:L/F)	0.611	0.469-0.798	<0.001	13.127	<0.0001
Males only	*fru* allele (S:L)	1.039	0.854-1.263	0.705	1.276	0.259
	Chro. (D:B)	1.301	1.061-1.595	0.011	3.119	0.077
	Allele x chro. (S/D:L/F)	0.772	0.58-1.029	0.077	3.117	0.077

Table 2.1. Results from Cox Proportional Hazard (CPH) models applied to lifespan data. Five models were used. One was for all flies and then the data was split to have separate models for each chromosome complement (B and D) and sex (female or male). The first column indicates the set of data that the model is applied to, while the second column indicates the term being tested in that model. CPH models use one level of a term as the reference level with a value of one. Other levels are then compared to this. The comparison made is shown in brackets as: (compared level:reference). Each term in a model has a hazard-ratio (H-R), a 95% confidence interval and a H-R p-value, which indicates if the compared level differs from the reference level. Also presented are χ_1^2 and its p-value, indicating the contribution of each term to the overall risk of mortality.

	B ♂	B ♀	D ♂	D ♀
Female fecundity	NA	S > L	NA	S > L
Male mating success	S = L	NA	S > L	NA
Larval survival	S < L	S < L	S < L	S < L
Development time	S = L	S = L	S = L	S = L
Lifespan	S = L	S < L	S = L	S = L

Table 2.2. Summary of the effects of *fru* alleles S and L on fitness components, in each sex and chromosome complement. The table indicates instances where, based on data, the S allele or the L allele resulted in greater or smaller values (S > L and S < L, respectively) or similar values (S = L) for measures of a fitness component. NA denotes cases where a trait could not be measured.

Chapter 3

3. Determining the mode of selection acting at the *Drosophila melanogaster fruitless* locus by tracking allele frequency trajectories under replicated experimental evolution

3.1 Declaration

All work presented here is my own, except for parts of section 3.4.2 ‘Cage population establishment and maintenance’ which was performed by Filip Ruzicka with help from Harvindar Pawar and Olivia Donaldson. Additionally, SLiM scripts for selection modelling were written in collaboration with Carl Mackintosh.

3.2 Abstract

The question of how genetic variation is maintained is a perennial problem in population genetics. One of the solutions to this problem is balancing selection. In Chapter 2 I showed that fitness patterns at the *fru* locus may generate balancing selection which had also been suggested by genomic signatures of variation. However, neither of these findings demonstrate that the *fru* locus is currently maintained by balancing selection. Evolve and re-sequence studies have proved a valuable tool to detect loci under selection, but have rarely been applied to balancing selection. I conducted experimental evolution on 10 cage populations of *Drosophila melanogaster* and performed pool-sequencing on flies from these populations. I constructed allele frequency trajectories for three proxy SNPs upstream of the *fruitless* (*fru*) polymorphism to track the frequency the *fru* polymorphism over 56 generations. I next used Approximate Bayesian Computation to diagnose the mode of selection at *fru* and differentiate between neutrality, directional or balancing selection. I found that all three proxy SNPs, and by extension most likely the *fru* indel polymorphism, were detected as under balancing selection. While in line with the findings of Chapter 2, these results are complicated by the fact that 44/100 putatively neutral short intron SNPs were also diagnosed as under balancing selection. It therefore remains unclear whether balancing selection at the *fru* locus reflects the fitness effects we detected at this locus, or other phenomena, such as associative overdominance. Future efforts should seek to clarify if this signature is real and could also extend this work to diagnose the mode of selection acting at loci within the *fru* cascade or those identified from association studies.

3.3 Introduction

A perennial problem in population genetics is why populations harbour large and stable amounts of genetic variation (Charlesworth and Hughes 2000; Lewontin 1974). Variation occurs due to mutation and is removed over time due to selection, as populations move towards fitness optima, or lost stochastically by drift (Leffler et al. 2012; Ellegren and Gaultier 2016; Muller 1950, 1958). Early investigations using allozyme data revealed abundant genetic variation in wild populations (Lewontin and Hubby 1966), which could be accounted for by the neutral (Kimura 1968) and subsequent nearly neutral theories (Ohta 1973, 1992) which proposed that most of the observed genetic variation had no or very little effect on fitness. The lack of any substantial evidence to the contrary led to the mutation-selection-drift balance theory becoming the dominant explanation for the maintenance of genetic variation (Beatty 1987; Charlesworth 2015; Prout 2000). The amount of genetic variation in a population was therefore considered to be a result of mutation rate or population size (Lewontin 1974; Leffler et al. 2012; Ellegren and Galtier 2016) with occasional non-neutral beneficial mutations being rapidly fixed by selection or rare cases of overdominance maintaining a few specific polymorphisms (Hedrick 2012, Kimura 1968). However, this consensus has been challenged recently, in part due to the rapid development of next-generation sequencing technologies (Casillas and Barbadilla 2017).

The comparison of genome sequences has led to some intriguing findings that have led to a re-examination of how genetic variation is maintained. The first comes from a study by Charlesworth (2015), which compared measures of genetic diversity in wild populations of *Drosophila melanogaster.* He found that the extent and patterns of genetic diversity were not compatible with maintenance by mutation-selection-drift balance alone and therefore concluded that genetic diversity must be maintained by another force: balancing selection. The second comes from the identification of loci

under long-term balancing selection (Fijarcsk and Babik 2015). Genomic scans compare sequences from multiple individuals and locate regions where the allele frequency is indicative of that expected under balancing selection (Charlesworth 2006). Andrés et al. (2009) found 60 genes in the human genome with elevated levels of genetic diversity compared to neutral expectations, with intermediate allele frequencies shared across long-separated populations. They concluded that this was due to long-term balancing selection maintaining genetic variation at these loci. This study was followed by others that have detected patterns of long-term balancing selection in humans (Bitarello et al. 2018; Siewert and Voight 2017), chimpanzees (Leffler et al. 2013), fish (Barson et al. 2015; Bernatchez 2016), insects (Croze et al. 2017; Lindtde et al. 2017; Sayadi et al. 2019; Unckless et al. 2016), pathogens (Amambua-Ngwa et al. 2012), and plants (Höger et al. 2012; Karasov et al. 2014; Koenig et al. 2019). In light of these studies, it appears that mutation-selection-drift balance is not the only force by which genetic diversity is maintained.

There are limitations as to what these studies can tell us, however. One problem is that a genomic signature of balancing selection doesn't inform us about the fitness effects resulting from such variation and how these give rise to a mechanism that generates balancing selection. Allele frequency patterns are also influenced by other factors, such as drift or demography, and so balancing selection is difficult to confirm from these alone. Phenotypic studies of polymorphisms have been useful in this regard where experiments are performed that assay the fitness effects of alleles at polymorphic loci to see if these are consistent with balancing selection. While these experiments can directly link variation and fitness effects, the identification of the causative genetic variation in the first place can be difficult, and fitness assays are extremely labour intensive. Due to this, very few examples of such polymorphisms have been described (see Smith et al. 2011; Mérot et al. 2020; Johnstone et al. 2013; Glaser-Schmitt et al. 2021).

In Chapter 2 I described the fitness effects of an indel polymorphism in the *D. melanogaster* gene *fruitless* (*fru*). Fitness assays showed that the S allele was associated with higher reproductive fitness in both sexes, while the L allele was associated with higher rates of larval survival to adulthood. Such a fitness pattern, where the allele with the highest fitness varies between different life-history traits, is indicative of antagonistic pleiotropy (AP), which is theoretically a mechanism for generating balancing selection. Therefore, I hypothesised that the *fru* polymorphism is maintained due to balancing selection occurring at this locus.

However, evidence of fitness variation is not the same as demonstrating the action of balancing selection itself. For this we need to show that the candidate polymorphism is actively maintained by balancing selection at the population level. This is something that neither genomic or phenotypic approaches can satisfactorily manage. To ultimately show that a polymorphism is maintained by balancing selection we must show: (i) there is a genomic signature of long-term maintenance; (ii) the polymorphism has variable effects on fitness consistent with the maintenance of that polymorphism; and (iii) the polymorphism is demonstrated to currently be maintained by balancing selection at the population level. Without a knowledge of all these complementary pieces of information, our ability to demonstrate unequivocally which loci are evolving under balancing selection and are thereby contributing to maintaining genetic variation is limited.

Our work with the *D. melanogaster* gene *fruitless* (*fru*) does provide an opportunity to accomplish this. Previous work identified that the indel polymorphism showed elevated levels of intermediate allele frequencies across distant populations (Hill 2017; Ruzicka 2018) – a genetic signature of balancing selection. As mentioned above, in Chapter 2 I showed that the alleles of the *fru* polymorphism affect fitness in a manner demonstrative of antagonistic pleiotropy – a mechanism that can generate balancing selection. The next step is to experimentally test if the antagonistic relationship at the

fru polymorphism can generate balancing selection at the population level. This would form a unique example of balancing selection, uniting genomic, phenotypic, and population level experimental approaches to investigate the maintenance of a polymorphic locus.

Balanced polymorphisms are expected to behave in a predictable and replicable manner where the allele frequencies are expected to converge around an equilibrium frequency determined by the relative fitness of the heterozygotes and homozygotes across different episodes of selection (such as juvenile survival and adult reproduction). Furthermore, this equilibrium frequency should be attained regardless of the initial allele frequency in the population (Fijarcsk and Babik 2015). This is in contrast to directional selection where the allele frequency will move in a consistent direction until the allele is either fixed or lost, and with neutrality where the allele frequency changes stochastically because of drift. By tracking the allele frequency over time, we can see if its behaviour is consistent with that expected by balancing selection or not.

This can be done with the use of evolve and re-sequence studies (E&R) (Burke et al. 2010; Schlötterer et al. 2015; Turner et al. 2011), an approach that combines experimental evolution with high-throughput sequencing to characterise the genetic composition of populations (Kawecki et al. 2014; Schlötterer et al. 2015). Experimental populations as a whole are subjected to a selective regime over many generations and individuals are sampled and sequenced at the start and end of the experiment, as well as sometimes at regular intervals throughout the experimental period (Schlötterer et al. 2015; Franssen et al. 2017). Sequencing of these samples enables changes in allele frequency to be detected and tracked, and allows for patterns of selection to be detected and candidate loci underpinning certain traits to be identified (Burke et al. 2010; Schötterer et al. 2015). For example, Remolina et al. (2014) identified 156 genes that showed signals of adaptive divergence when selecting for long lived *D. melanogaster*. E&R

experiments typically use several replicate populations, which can involve sets of populations under opposite selective conditions, or populations with alternate starting mixtures of genetic variation. These replicates demonstrate whether patterns of allele frequency change are consistent. In the case of *fru*, the starting frequencies of the two alleles were manipulated and then tracked over time in otherwise identical populations. If the allele trajectory is consistent across cages regardless of the initial frequency of *fru*, showing signs of coalescing, and the polymorphism is maintained in all populations, this would show support for balancing selection. While E&R studies have huge potential, the approach is still relatively new and has not yet been applied to specifically look for balancing selection.

An approach related to E&R experiments has detected evidence of balancing selection by regularly sampling wild populations of *D. melanogaster* to track seasonal changes in allele frequencies. By correlating changes in allele frequencies with variation in conditions over space and time, Bergland et al. (2014) showed that allele frequencies of loci associated with adaptation to particular seasonal conditions vary across the year. An allele which confers better survival during the winter is selected for during that season, but selected against during the summer when an alternative allele is favoured (Bergland et al. 2014). Further sampling has shown that this fluctuating selection pattern occurs across clines as different populations are affected by similar patterns of seasonal variation, with consistency across years, metapopulations and continents in *D. melanogaster* (Kapun et al. 2020; Machondo et al. 2021) and other species (Han et al. 2020). Such patterns indicate the occurrence of balancing selection acting to maintain genetic diversity in traits related to seasonal adaptation. Both these studies and some E&R experiments make use of simulations to discern between changes in allele frequency that are due to balancing selection and those from other forms of selection (Wittman et al. 2017; Franssen et a. 2017). Selection modelling can simulate the fate of a locus under different modes of selection. These simulations can then be compared to the observed allele

frequency patterns to determine if these show signs of selection. The use of simulations can also help account for factors affecting the trajectory of the allele frequencies such as drift, the influence of linkage caused by inversions and regions of low recombination (Franssen et al. 2015) or the influence of linked deleterious mutations (Kapun et al. 2014).

E&R and population sampling studies frequently employ pool-sequencing (pool-seq) to obtain allele frequency measures. Pool-seq is a highly efficient method of collecting genomic data by sequencing whole groups of individuals, rather than each individual separately (Futschik and Schlötterer 2010; Schlötterer et al. 2014). Though some data is lost since individual genotypes are unknown, pool-seq is highly cost effective and produces high quality measures of population allele frequencies (Schlötterer et al. 2014; Gautier et al. 2013). Since pool-seq is cheaper than individual sequencing, lots of individuals can contribute to pools and data can be collected on a population as a whole, on multiple populations, or on the same population at multiple time points. (Bergland et al. 2014; Franssen et al. 2017; Schlötterer et al. 2015). Taking multiple pooled samples of the same population allows for the construction of allele trajectories, which follow the frequency of alleles over time (Franssen et al. 2017). This has been used to track a polymorphisms to see if they show signs of balancing selection (Bergland et al. 2014; Rudman et al. 2021).

In this chapter, I apply an approach inspired by E&R and population tracking studies to analyse the frequency trajectories of the two *fru* alleles over time in replicate populations of *D. melanogaster* and diagnose the mode of selection acting at the *fru* locus. I used a series of 10 populations of *D. melanogaster* evolving in parallel, with biased starting frequencies of the *fru* alleles of either 0.9 or 0.1, to detect changes in allele frequency indicative of balancing selection. I refer to this study as an ‘experiment’, but should perhaps more truly be classed as ‘laboratory natural selection’ due to a lack of a true control. However, previous studies using a similar design have used

the term experimental evolution (Schlötterer et al. 2015) and I follow that example here. Populations were sampled periodically across >50 generations. I used pool-seq to generate whole genome sequencing data for these populations at nine timepoints, from which I created time series trajectories of *fru* allele frequencies.

I analysed this data using Approximate Bayesian Computation (ABC), a powerful method based on contrasting observed data with large numbers of datasets generated by simulation under different modes of selection and a wide range of parameter values. The approach can then be used to assess the (posterior) probability of different models and infer the most likely parameter values (Csilléry et al. 2010). ABC was also used to determine the selection parameters responsible for the behaviour of the *fru* locus in our populations. Cross-referencing the results obtained for *fru* to putatively neutral loci located in short introns provides a method to compare our results to neutral expectations and enabled me to assess potentially confounding factors such as associative overdominance (Clemente and Vogl 2012; Parsch et al. 2010; Schlötterer et al. 2015). Through this process it appears as though the *fru* locus is under some form of balancing selection in our populations. However, there is also balancing selection detected at 44 out of 100 putatively neutral short intron SNPs that were analysed in parallel, which casts doubt on whether the patterns of selection at *fru* are truly due to balancing selection or other forces such as selection at linked loci.

3.4 Methods

3.4.1 Fly culture and husbandry

All flies were maintained at 25°C in constant temperature rooms at 50% humidity on a 12:12hr light-dark cycle. When required, flies were anaesthetised using a CO_2 pad for short periods of time and manipulated using an aspirator or paint brush. Food was supplied as corn-molasses media either in vials (8ml) or bottles (50ml).

3.4.2 Cage population establishment and maintenance

In order to be able to manipulate the starting frequency of the *fru* allele in each of the experimental populations, two populations were created, one fixed for the S and the other for L allele, but variable elsewhere across the genome. Each population was composed of 14 inbred lines, selected from the *Drosophila* Genetics Reference Panel, or DGRP, a set of 205 fully inbred lines collected from a wild population in Raleigh, North Carolina (Mackay et al. 2012). Each of these lines has been sequenced and the genotype at genome-wide SNPs is known (Huang et al. 2014). The specific lines used to establish each population are listed in Table 3.1. The 14 lines for each population were mated together in a round robin design (♀1 x ♂2, ♀2 x ♂3, ♀3 x ♂4, etc.) using 25 virgin offspring from each line. The resulting offspring were then mixed together and equally allocated across three bottles for 3 generations.

To track the population allele frequency of the *fru* polymorphism through time, 10 replicate populations of *D. melanogaster* were established by selecting individuals from these two populations. Five were established with a starting allele ratio of 9:1 S:L and the other five established with a 1:9 S:L ratio. For the five cages starting with a high frequency of the L allele, 450 virgin flies (225 males, 225 females) were chosen per cage from the population fixed for the L allele, and 50 virgin flies (25 males, 25 females) per

cage chosen from the population fixed for the S allele. This results in a population with a *fru* allele proportion of nine L alleles to every 1 S allele, in other words, an S allele frequency of 0.1. The same process was repeated for those populations which started with a high S allele frequency, but on this occasion taking the higher numbers of flies from the population fixed for the S allele. These cages have a starting S allele frequency of 0.9. At this stage, each cage was provided with three bottles of media, sprinkled with additional yeast. Each week three bottles were added to each cage until each cage contained 12 bottles. After this, the three oldest bottles of media were removed each week and replaced with three new bottles, ensuring that each cage always contained 12 bottles of media.

Each population was housed in a cage made of hard transparent plastic, 33cm x 22cm x 20cm, with two circular holes (12cmø) at each end. One of these holes was covered with a piece of cloth, glued around the edge, which provided air flow. The other hole had a lip of plastic around the edge and over this was fitted a cut piece of black tights, approximately 20cm long. One end of this material was secured around the plastic lip with an elastic band, while the other end was knotted to prevent the flies escaping. This could be easily untied to allow access to the cage for food replacement and periodic sampling of flies. See Figure 3.1 for photos of an example cage.

3.4.3 Cage sampling and pool-sequencing

Experimental cage populations were established in July 2017. Approximately 100 flies were sampled using an aspirator from each cage each month for the first two months of the experiment and every second month after that. Fly samples were checked for an roughly equal sex ratio and frozen at -20°C.

Due to their large number, it was not feasible to sequence all samples collected over the course of the experiment, even when using pool-seq. A decision therefore had to made regarding the experimental design for the sequencing, including the number of timepoints to sequence, the number of

flies for each population and sample to be included in the sequenced DNA pool, and the sequencing coverage for each pool. The first decision was to include 48 flies (or 96 chromosomes) per pool. This was the highest number of flies that could be included while keeping the number of flies in each pool the same because some flies had been previously used for individual PCR genotyping. It is also greater than the minimum number of flies advised for pooled sequencing by Schlötterer et al. (2014).

To guide the decision on the number of timepoints and depth of sequencing, a series of simulations were performed. I produced a simulated data set of a population where an allele, S, increases in frequency linearly from 0.1 to 0.7 over 12 equally spaced timepoints. I then simulated subjecting this population to pool-sequencing by performing 2 rounds of binomial sampling (chromosomes sampled from the population and reads sampled from the chromosome pool). The total volume of sequencing data generated (and therefore the amount of money spent) under each scenario was fixed. I then plotted the results to see how well different combinations of coverage and timepoints affected the estimation of the increase of the S allele over time (Figure 3.2). Combining this information with published best practices for pool-seq experiments (Schlötterer et al. 2014), the decision was made to sequence flies from nine different timepoints for a total of 90 pool samples, at a coverage of 40X. It was also decided to spread timepoints in an uneven manner since most change was expected to occur during the first few months of the experiment. The nine collections that were chosen were August, September, and November 2017; January, May, and September 2018; and January, May and November 2019. Assuming a generation time of two weeks, these were 2, 4, 8, 12, 20, 28, 36, 44 and 56 generations from the start of the experiment.

To avoid contamination from gut microbiota, sperm, and fertilised eggs, DNA was extracted from fly heads (Griffin et al. 2017). The heads were removed by briefly freezing an Eppendorf tube containing a group of flies in

liquid nitrogen. The tube was then held against a vortex to spin the flies inside, detaching the flies' heads from their bodies. The contents were then transferred to a glass petri dish and the heads sifted out using a paint brush and placed together in a new tube. To try and ensure that all heads contributed to the pool of DNA equally, heads were placed in groups of eight with six groups per pool.

DNA was extracted from each group of eight heads using a DNeasy Blood and Tissue kit (Qiagen, Velno, Netherlands). DNA concentration was measured for each group of eight and the six groups combined such that each contributed the same amount of DNA to the overall DNA pool from the 48 fly heads. This was repeated for each sample to create 90 pools. After checking DNA concentration and quality, the samples were shipped on dry ice for library preparation and sequencing at Novogene (UK) Company Limited. DNA libraries were prepared using NEBNext DNA Library Prep Kit and 1µg of DNA per sample. Resulting PCR products were purified (AMPure XP system) and the resulting libraries analysed for size distribution using Agilent 2100 Bioanalyzer and quantified using real-time PCR. Whole genome sequencing for all 90 samples was performed using Illumina paired-end 150bp reads.

3.4.4 Bioinformatics and data analysis

Data was received from Novogene in a fastq file format. Adapter sequences were removed and reads trimmed with Cutadapt (version 2.5, Martin 2011). Reads were mapped to the *D. melanogaster* reference genome release 6 (Hoskins et al. 2015; dos Santos et al. 2015) using bowtie 2 (version 2.3.5.2, Langmead and Salzberg 2012) producing a SAM output file. Samtools (version 1.9; Li et al. 2009) was used to remove poorly mapped reads (quality <20), sort SAM files and convert to BAM files. Duplicate reads were removed and read groups defined with Picard-tools (version 2.20.3; Picard Toolkit, 2019), before indexing BAM files with Samtools. Indel realignment was then carried out using the 'RealignerTargetCreator' and then

'IndelRealigner' tools in GATK (version 3.8.1.0; McKenna et al. 2010; Van der Auwera et al. 2013). Once all 90 samples were processed, BAM files were indexed and data for genomic positions of interest combined into a single mpileup file using Samtools. I focussed on two sets of positions, a ~131kbp region around the *fru* indel and a set of SNPs in short autosomal introns that served as an a priori neutral point of comparison. Variant calls for the selected positions were produced by the package Popoolation2 (Kofler et al. 2011) in a two-step process. First, the mpileup file was converted into a sync file which records the number of each nucleotide at each position for every sample. This file was then converted to an _rc file which contains a count of the number of each variant for each pool sample for polymorphic sites only.

All further analysis was performed in RStudio (R core Team, 2019) and figures produced using the package ggplot2 (Wickham, 2016). The analysis included the following steps, summarised in Figure 3.3. The _rc file was filtered to include biallelic SNPs and allele counts converted to frequencies of the same allele. The fact that the *fru* polymorphism is an indel, an insertion relative to the reference genome (Ruzicka 2018, flybase.org 2021a), compromises read mapping in the immediate vicinity of the indel and can therefore result in a read desert (Bennet et al. 2020; Palmieri and Schötterer 2009; Ratan et al. 2015). To track the frequency of the indel, I therefore needed to identify SNPs in linkage with the indel, but far enough away that they had sufficient read coverage. I first refined the list of SNPs to a smaller 9.7kbp region from positions 18,516,272 to 18,525,973 on chromosome arm 3R, resulting in 4001 SNPs. Since many of these SNPs are not real polymorphisms but are isolated alternative nucleotide calls due to sequencing error, I developed a polymorphism filter to select SNPs that displayed significant levels of polymorphism similar to the *fru* locus. This focussed solely on samples from timepoint 1 which was estimated to be 2 generations after the start of the experiment. Since the 10 cages started at a frequency of either 0.9 or 0.1 for each *fru* allele, any true polymorphism

linked to the indel should show appreciable frequency variation among the cages. The filter therefore selects those SNPs where the allele that was more common across all cages was found at a frequency of <0.5 in at least 4/10 cages, i.e. was the less common allele in those cages. This criterion produced a much smaller list of 19 possible candidate SNPs. Three of these occurred within 525bp upstream of the *fru* indel. PLINK (Purcell et al. 2007) was used to calculate linkage disequilibrium between the locus nearest to the indel included in the DGRP dataset (obtained from http://dgrp2.gnets.ncsu.edu/data.html) (18,520,985) and loci closest to the proxy SNPs. All further analyses focus on these three SNPs as replicate proxies for the *fru* indel.

3.4.5 Diagnosing the mode of selection acting at the *fru* polymorphism using Approximate Bayesian Computation

To diagnose the form of selection acting at the *fru* locus, Approximate Bayesian Computation (ABC) analysis was performed using the package 'abc' in R (Csillery et al. 2012). ABC is a statistical method which uses simulated data to describe observed data patterns that would be difficult to analyse using traditional likelihood methods. ABC works by comparing observed data to simulations which describe an experiment or biological phenomenon of interest under one or more models and a range of parameter values. ABC can then be applied to investigate two aspects of the data: 1) 'model choice' - if several models are considered, the ABC can determine which is the most probable given the observed data, and 2) 'parameter estimation' - which parameter values most closely recreate the observed data. ABC makes these judgements by assessing the similarity between observed data and simulations, based on a number of 'summary statistics', which are calculated for both the observed and simulated data and describe some aspect of their properties. Those simulations where the summary statistics most closely match the data are retained and used to calculate posterior probabilities of alternative models (e.g. the proportional representation of different models among the retained set) and posterior

distributions of parameter values (parameter values used in simulations in the retained set). The size of the posterior distribution is determined by the tolerance rate, a percentage of the total dataset to be included, while the matching of simulated and observed statistics can be performed using different algorithms. The best algorithm will vary depending on the number, complexity and covariance of the summary statistics and parameter values, and should be determined by cross-validation before the main ABC analysis. I used ABC for both 'model choice', to diagnose the mode of selection acting at on the three proxy *fru* SNPs from a choice of five alternative selection models, and 'parameter estimation', to estimate the selection parameters which most accurately describe the observed data.

3.4.5.1 Population modelling using SLiM

Simulations were performed in SLiM 3 (Haller and Messer 2019). I implemented a standard Wright-Fisher model with a single locus with two alleles (the L and S alleles). Simulations were designed to mimic the structure of the actual study as closely as possible. They included ten populations with no migration, five starting with a frequency of 0.9 of the focal allele and the other five with a frequency of 0.1. Simulations were run at a fixed population size of 1000 (also meaning N_e=1000 in the Wright-Fisher model). The real-life cage populations started with 500 flies established from 28 DGRP lines, but quickly grew to a size of several thousand. An N_e of 1000 was used to accommodate the fact that effective size is typically significantly smaller than the census size of a population, due to factors such as elevated reproductive variance (Mueller et al. 2013; Phillips et al. 2016; Wright 1931). In the first generation, chromosomes in each population were randomly seeded with the focal and alternative alleles according to the corresponding starting frequencies of the population. Simulations were then run for 56 generations and the frequency of the focal allele tracked.

Selection was implemented using a selection coefficient, *s*, and dominance coefficient, *h*, associated with the focal allele. Simulations were

run for five selection models: *neutrality*, *positive* and *negative directional selection*, and *positive* and *negative balancing selection*.. Under *neutrality*, s was set to 0, while under *positive directional selection* $s>0$ and $0<h<1$ and under *negative directional selection* $s<0$ and $0<h<1$. Models of balancing selection were implemented as overdominance, where the heterozygote genotype is fitter than either homozygote. In Chapter 2 I state that the mechanism I believe is acting at the *fru* locus is AP rather than overdominance. Distinguishing between the two mechanisms of balancing selection can be difficult and depends on multiple fitness traits being considered. The selection models implemented here do not distinguish between different traits, but summarise fitness as whole. This is both simpler mathematically to describe and allows direct comparison with other selection models by using the same two parameters. Additionally, the behaviour of loci under overdominance and AP is similar regardless of the underlying mechanisms. For *positive balancing selection* we set $s>0$ and $h>1$, resulting in the focal homozygote being fitter than the alternative homozygote, but less fit than the heterozygote. To accommodate the opposite case of *negative balancing selection*, where a focal allele homozygote has the lowest fitness, I set selection parameters to $s'>0$ and $h'>1$ for the alternative allele and then converted them as $s = -s'/(1 + s')$ and $h = 1 - h'$ for the focal allele. For simulations with selection, the values of s and h were drawn from a uniform distribution within parameter ranges provided in Table 3.2, following pilot analyses (see Appendix C for details).

One million simulation runs were performed for each selection model. In each, the frequency of the focal allele in each population was recorded at the end of generations 2, 4, 8, 12, 20, 28, 36, 44 and 56, in line with the sampling times of the cages. As these data were true frequencies, determined by counting alleles among all chromosomes in a population, two further rounds of binomial sampling were applied to mimic the actual experiment. The first was the selection of 96 chromosomes (48 flies) from each population to represent the cage sampling at each timepoint. This used

the simulated true frequencies for each timepoint and population as binomial probabilities. The second round of binomial sampling represented the number of reads (coverage) for a SNP at a particular timepoint and for a specific population. This second sampling was performed separately for each of the three empirically observed *fru* SNPs and used the frequencies obtained in the first round of binomial sampling as probabilities and the observed read depth of the 90 empirical samples for a particular SNP as the numbers of binomial draws.

3.4.5.2 Summary statistics for describing frequency trajectories

The frequency values at each timepoint were used to construct allele frequency trajectories for both the proxy SNPs and simulations which describe the allele frequency pattern over time. To capture the variation in these patterns, I calculated 13 summary statistics that describe the behaviour of a focal allele across all cages over the course of the experiment, where the focal allele is arbitrarily chosen to be the SNP allele associated with the S form for the *fru* indel polymorphism. Each of these summary statistics were designed to help distinguish between different selection models, for example SNPs under directional selection will more likely fix than those under balancing selection. The information contained in the statistics was assessed using correlation matrices and violin plots (see Appendix C for deatils). The 13 summary statistics used in this analysis are:

1. Fixation – the proportion of cages where fixation occurred. Frequency must be 0 or 1 in the last sampled generation and 0 or 1 in the second to last generation to confirmed as fixed.
2. Persistence time – if fixation occurs, the time in generations that it took to fix. Calculated as the earliest generation where the frequency is 0 or 1 and all succeeding frequencies are 0 or 1.
3. Absolute change – the final frequency per cage minus the starting frequency of that cage (0. 9 or 0.1), converted to absolute value. A mean is taken across all cages.

4. Slope of best fit line – a best fit line is fitted to the frequency trajectory for each cage, the gradient calculated, and a mean taken of the 10 cages.
5. Sum of squares – the sum of the squared residual values from the best fit line. A mean is taken of the 10 cages.
6. Monotonicity – a measure of the degree of monotony in the frequency trajectory. Starting from 0, the score of this metric is incremented by +1 or -1 each time the frequency change from one timepoint to the next has an absolute value of >0.05 or <-0.05, respectively. Scores are then averaged across cages.
7. Final variance – the variance of the frequency values at generation 56.
8. Change in variance – the variance at generation 56 minus the variance in frequency at the first sampled timepoint (generation 2).
9. The mean number of timepoints close to the final frequency – count how many of the previous timepoints have a frequency <0.1 different from the final frequency value. A mean is taken of the 10 cages.
10. Mean change per generation – the mean change in frequency per generation. Calculated per cage and then averaged across all cages.
11. Final values within 0.1 range – a mean of the frequency at the final timepoint (generation 56) is calculated. Then the proportion of cages within 0.1 of this value is calculated.
12. Final mean frequency – mean of the final frequency value of each cage.
13. Difference in final value means – the mean final frequency value of the five cages which started at a low frequency (0.1) is subtracted from the mean final value of the five cages starting with at high frequency (0.9).

3.4.5.3 Application of ABC

With summary statistics calculated for all simulations models and the three proxy SNPs, ABC was applied to diagnose the mode of selection occurring at the *fru* locus using the model choice function. First, I performed a cross-validation step to ensure that the different modes of selection could be differentiated from each other using different matching algorithms. This randomly took 50 simulations from each class of selection model and asked

the program if it could correctly identify which class of selection model it belonged to base only on its summary statistic values. This cross-validation showed that the use of the multinomial logistic regression algorithm could accurately discern between all five modes of selection (Figure 3.4).

Model choice was then performed to diagnose the mode of the selection the proxy SNPs were under. This is done by calculating the proportion of simulations from each selection model class which contribute to the posterior distribution as formed by ABC. This used a 5% tolerance limit and the multinomial logistic regression algorithm. I combined models into three types: neutral, directional and balancing and diagnosed selection if the posterior probability was >0.5.

The final step was to perform model goodness-of-fit. The previous model choice step only identifies which of the alternative models has the highest probability of having generated the data. This doesn't necessarily mean that it is an ideal model, just that it is better than the alternatives. Model goodness-of-fit checks if the simulated data overlaps sufficiently with the real data to see if it could have feasibly produced that data. This was done using 100 samples from each selection model.

Parameter estimation was also performed using the 'abc' package (Csillery et al. 2012). Cross validation of parameter estimation was first done to compare the error in estimating *s* and *h* associated with three estimation algorithms (Appendix C). Parameter estimation was then performed using the 'neural net' algorithms with a 0.5% tolerance to estimate *s* and *h* for each SNP.

3.4.6 Short intron SNPs

In order to provide a set of control SNPs as comparison for those associated with the *fru* polymorphism, I also analysed data for 100 autosomal SNPs located in short introns. Short intron polymorphisms have been found to be

close to neutral in *D. melanogaster* (Clemente and Vogl 2012). The allele frequency at these loci at the beginning of the experiment was estimated based on the genome sequences of the DGRP lines used to establish the cage populations. For this purpose, I filtered the VCF file for all DGRP lines (obtained from http://dgrp2.gnets.ncsu.edu/data.html) using vcftools (version 0.1.16; Danecek et al. 2011) to create two subset files, one containing data for the 14 lines used to establish the population fixed for the S allele and another for the 14 lines used to establish the population fixed for the L allele (Table 3.1). Starting from a list of 11,142 short intron SNPs identified among genome sequences from worldwide *D. melanogaster* populations (Lack et al. 2015, 2016) I identified sites that were polymorphic in the sets of the DGRP lines used to found the cage populations. For each of those SNPs I then calculated the allele frequencies in each set and then combined those in a weighted average to obtain the expected starting frequencies in the cages starting at 9:1 and 1:9 S:L ratio. This approach assumed that all lines contributed equally to the establishment of the S and L fixed populations. The list of short intron SNPs was filtered to remove SNPs where the expected allele frequency was either <0.025 or >0.975, as polymorphism at these loci could be easily lost due to drift or unequal sampling.

Popoolation2 was used to create the _rc files for these filtered SNPs in the same way as above (Figure 3.3). The _rc file was then used to filter for only biallelic short intron SNPs. From the resulting list, 100 loci widely spaced along the major autosomal chromosome arms (2L, 2R, 3L and 3R) were chosen to give a semi-random spread of putatively neutral sites from across the genome. Frequency data from these SNPs were treated in the same manner as those from the *fru* SNPs, including calculating summary statistics from the observed frequency trajectories and ABC model choice and parameter inference. SLiM simulations were started from the frequencies calculated for each SNP in the two types of cages (see above).

Based on the ABC results, SNPs were assigned to a particular mode of selection if the posterior probability of that model was greater than 0.5 and results were summarised as counts of SNPs associated with each mode of selection. An additional count combined the two balancing selection models and the two directional selection models. Parameter estimation was performed using a 5% tolerance, therefore producing a posterior distribution for each parameter of 25,000 values (from a total of 5,000,000 simulations) for each locus. Parameter values were converted to positive values of the fitter allele to make these easier to compare.

Previous experimental evolution studies using *D. melanogaster* similar to this one, have noted issues in assigning the effect of selection at a SNP due to its position within large segregating inversions (Kapun et al. 2014). These inversions are frequent in *D. melanogaster* and have been linked to adaptive clines and adaptation (Durmaz et al. 2018; Kapun et al. 2016; Kapun and Flatt 2019). Since inversions do not recombine with equivalent chromosomal regions, linkage disequilibrium is elevated and the effect of linked selection can be much larger within these inversions, meaning that neutral sites can be affected by hitchhiking and background selection due to selection on linked loci within the inversion. Two large inversions are carried by the 28 DGRP lines used to found our populations In(2L)t (2L: 2,225,744–13,154,180, 10.93Mbp) and In(3R)Mo (3R: 21,406,917–29,031,297, 7.62Mbp) (data from dgrp2.gnets.ncsu.edu/data.html). To account for the potential hitchhiking effect caused by these inversions I checked for short intron loci occurring within the range of the two inversions. I then counted the number of loci within each inversion to see if there were any biases in the type of selection model they had reported the greatest probability for. Additionally, another study by Franssen et al. (2015) identified regions in the *D. melanogaster* genome with particularly low recombination rates which could have similar effects on linked loci as inversions do. I therefore checked if any short-intron SNPs occur in these regions in order to check for bias in the mode of selection.

3.5 Results

By November 2019, 28 months after the start of the experiment, flies had been collected at 15 timepoints (every 2 months plus August 2017) resulting in ~15,000 flies sampled. All cages supported viable breeding populations, with several thousand flies in each. Although the sides of cages had become obscured by fly excrement, the flies appeared healthy and no outbreaks of mould or other contamination issues occurred in any cage.

3.5.1 Polymorphism around the *fru* indel

The total volume of sequencing data received from Novogene was 696.5Gbp across all 90 samples. Each sample had between 42,260,026 and 69,790,844 million reads. The sequencing error rate was estimated at 0.03%.

A 9,701bp region of chromosome 3R, centred around the position of the *fru* indel, was specified containing 4,001 SNPs. Within this, 19 SNPs met the polymorphism criterion where at least 4 cages had an allele frequency of the most common allele across the ten cages less than 0.5. Of these, three SNPs were chosen as proxies for the *fru* indel polymorphism. They were the three closest SNPs to the *fru* indel and their frequencies showed the greatest difference between the two sets of cages and were in line with expectations from the 9:1 and 1:9 ratios in which the cages were set up. These proxy SNPs were located at positions 18,521,181, 18,521,208, and 18,521,521 on chromosome arm 3R. In all further analysis these are referred to by the last 4 digits of their chromosomal position, i.e. 1181, 1208, 1521. These loci were 183, 210, and 523bp upstream of the *fru* indel located at 3R:18,520,998 (*fru* is on the reverse strand) and all are within the same regulatory region as the *fru* polymorphism which is located between 3R:18,520,191 and 3R:18,523,122 (flybase.org 2021b). SNPs 1181, 1208, and 1521 had respective sequencing coverages of 21.3, 24.1, and 40.0. Because none of the proxy SNPs were covered in the polymorphism data available for the DGRP lines, linkage disequilibrium was calculated between the loci in those

data that were closest to the proxy SNPs (position 18,521,175, which lies 6 and 28bp downstream of proxy SNPs 1181 and 1208, respectively, and position 18,521,462, which lies 59bp downstream of 1521) and that closest to the indel (position 18,520,985, which lies 13bp downstream for the indel). R^2 between positions 18,521,175 and 18,520,985 was 0.86, while that between 18,521,462 and 18,520,985 was 0.4.

3.5.2 Investigation of *fru* using Approximate Bayesian Computation

Frequency trajectories of the three proxy SNPs are shown in Figure 3.5 and the values of the summary statistics are presented in Table 3.3. The patterns of change suggest a convergence of frequencies towards intermediate values over the course of the experiment. In line with this impression, performing ABC model choice on the empirical data provided strong support for balancing selection for all three proxy SNPs based on their posterior probabilities (1181: post. prob.=0.815, 1208: post. prob.=0. 985, 1521: post. prob.=0. 999). No other mode of selection achieved a probability >0.2. The maximum probability value of any model can have is 1. Full model choice results are presented in Table 3.4. Goodness-of-fit testing showed a good fit (p>0.05) between the individual model with the strongest support, *positive balancing selection*, and the observed data of all three SNPs (1181: D=8.3, p=0.13; 1208: D=10.4, p=0.07; 1521: D=10.32, p=0.11). Full goodness-of-fit results are presented in Table 3.5.

Parameter estimation with neural net and a posterior distribution size of 20,000 produced an estimated mean *s* between 0.01 and 0.017 (1181: mean *s*=0.0474, 95%HPD=0.005–0.056; 1208: mean *s*=0.01, 95%HPD=0.0004–0.026; 1521: mean *s*=0.017, 95%HPD=0.001–0.04). A mean value of *h* between 1.721 and 3.797 was estimated for the three SNPs (1181: mean *h*=2.032, 95%CI=0.932–3.819; 1208: mean *h*=1.721, 95%CI=0.126–5.144; 1521: mean *h*=3.797, 95%CI=2.578–5.074). From these estimated parameter values equilibrium frequencies of the minor allele (MAF) of 0.337, 0.295, and 0.422 were calculated for SNPs 1181, 1208 and

1521 respectively. Posterior distributions of estimated parameters are shown in Figure 3.6.

3.5.3 Short intron SNPs

Of the 100 short-intron loci, combining selection types resulted in 44 SNPs being diagnosed as under balancing selection, 37 as under directional selection, 13 as evolving neutrally, and 6 as undefined (all posterior probabilities <0.5) (Figure 3.7 and Table 3.6).

Parameter estimates were converted to be those of the favoured allele for easier comparison between SNPs. The median mean *s* was 0.102 and estimates ranged from 0.004 to 0.253. These values extend beyond the range of the input parameters as the neural net is capable of projecting beyond these limits. Posterior means of *h* had a median of 1.008 and the values ranged between 0.005 and 3.975. These values were used to calculate predicted equilibrium frequencies for the minor allele are shown in Figure 3.8. Comparing parameter estimate values to those produced for the three proxy SNPs, the three SNPs show a distinct combination of high *h* and low *s* compared to the short intron loci (Figure 3.9).

Investigation of short introns loci located within large inversions found 17 loci within the In(2L)t inversion (2L: 2,225,744 – 13,154,180). Nine loci were diagnosed as balancing selection (all positive), three as directional selection (all positive), four as neutral, and for one locus no model had a probability >0.5. Of seven loci located within inversion In(3R)Mo (3R: 21,406,917 – 29,031,297), five loci were diagnosed as directional selection (1 positive and 4 negative) and two loci as *positive balancing selection*. The patterns of selection of these SNPs did not differ from that overall (χ_3^2=1.272, p=0.736) Regions of low recombination in the *D. melanogaster* genome identified by Franssen et al. (2015) covered 15,112,165bp. 12 short intron SNPs were located within these regions. Six were diagnosed as directional selection (three negative and three positive), 4 loci as balancing selection

(three positive and one negative), and two loci as neutral. This did not differ from the overall pattern (χ_3^2=1.702, p=0.636).

3.6 Discussion

3.6.1 Diagnosing selection at the *fru* polymorphism

In this chapter I used laboratory evolution to experimentally test if the *fru* indel polymorphism described in Chapter 2 is under balancing selection. By comparing allele frequency trajectories obtained from sequencing data of flies from replicate populations of *D. melanogaster* to those generated from simulations of five different selection models, it appears that the *fru* locus is under some form of balancing selection. Balancing selection, with a positive *s* for the S allele, is the most supported selection model, reporting a posterior probability of >0.5 for all three proxy SNPs (Table 3.4) when analysed using ABC. Thus, simulation runs produced under balancing selection models (positive or negative) produced allele frequency trajectories with summary statistics most similar to those of the empirical data from all three proxy SNPs. When the probabilities of the two balancing selection models are combined for SNPs 1181, 1208, and 1521, the overall probability of balancing selection is 0.816, 0.985, and 0.999 respectively. Therefore, balancing selection is clearly the most probable form of selection occurring at all three proxy SNPs. The *positive balancing selection* model also reported an agreeable goodness-of-fit with the data (Table 3.5), confirming that this mode of selection could plausibly have produced the summary statistics observed rather than simply being more probable than the alternatives. Since the three SNPs are in close linkage with the *fru* polymorphism, it would appear highly likely from these analyses that the *fru* indel polymorphism is also under balancing selection.

The evidence discussed thus far is consistent with the findings from Chapter 2. There I found a difference in the fitness benefits of the two *fru* alleles which was indicative of AP. AP has been proposed as a mechanism capable of generating balancing selection by which both alleles are maintained in the population. This is illustrated in Figure 3.5 where polymorphism is maintained in all 10 cage populations for each proxy SNP

and the trajectories show consistent change from their initial biased frequencies. This is further supported by the estimated equilibrium frequencies for the S allele. The use of experimental evolution to track alleles through time and diagnosis of balancing selection occurring is one of the first of its kind. This complements the fitness assay results from Chapter 2 and the previous work which found long-term patterns of balancing selection around the *fru* indel polymorphism (Hill 2017; Ruzicka 2018). These results in isolation support the hypothesis that the *fru* locus is under balancing selection which maintains genetic variation in real time.

3.6.2 Short intron loci

But in order to put the results observed for the *fru* locus in perspective, a comparison between the proxy SNPs and the genome at large is required. The most appropriate contrast is with neutral genetic variation which is unaffected by selection (Parsch et al. 2010). Very little of the *D. melanogaster* genome is truly neutral (Andolfatto 2005). However, loci located within short introns are considered a good neutral approximation (Parsch et al. 2010). The process of diagnosing the mode of selection acting was therefore applied to a data set of 100 short intron SNPs.

Of these 100 short intron SNPs, 13 were diagnosed as being neutral based on their model choice posterior probabilities (Table 3.6). A further six SNPs did not record posterior probabilities >0.5 for any mode of selection and no mode of selection could be diagnosed. This means that only between 18% of short intron loci are not under some form of selection. While it is encouraging that neutrality was detected for some of the short intron SNPs, the fact that the number is much lower than the number of loci diagnosed as under balancing (44) or directional selection (37) is worrying. Equilibrium frequencies based on estimated parameter values increase the number of loci expected to eventually fix, including some diagnosed as balancing selection (Figure 3.8).

Plotting the estimated *s* and *h* values shows that the three proxy SNPs are differentiated from the short intron loci with a unique combination of relatively low *s* and high *h* parameter estimates (Figure 3.9). These values show that the patterns near the *fru* polymorphism are at least different from the short intron loci. Nevertheless, the results of selection at the short intron SNPs are a potentially confounding result for this study. These SNPs may not be neutral in the purest sense and to have found a few with signatures of selection would have been unsurprising, but the fact that 82/100 were diagnosed as being under some form of selection, and the high incidence of balancing selection (44/100) in particular, are concerning. Ultimately the diagnosis of balancing selection at the *fru* locus, although stronger and with an unusual combination of selection parameter estimates, could be questioned if the high number of short intron SNPs diagnosed as balancing selection cannot be accounted for.

The first possible explanation is that balancing selection is much more pervasive in the *D. melanogaster* genome than previously thought. Previous studies have found signatures of long-term balancing selection in the *D. melanogaster* genome (Croze et al. 2017, Unckless et al. 2016; Sato et al. 2016), however, these are neither common nor widespread enough to account for the selection patterns of the short intron SNPs. It is possible that balancing selection resulting in younger polymorphisms than can be detected by these studies may be common in *D. melanogaster* and these may account for the patterns seen in short intron SNPs. While this scenario is possible, the discrepancy between the small number loci detected as under balancing selection in previous genome scans and the high incidence of balancing selection detected does not suggest this as a plausible explanation.

Secondly, results of previous studies aiming to detect loci under selection have been confounded by the presence of large segregating inversions in the *D. melanogaster* genome and by regions of low recombination (Franssen et al. 2015), which both prevent recombination and

thereby extend the distance over which linkage can occur (Kapun et al. 2014). However, the mode of selection diagnosed for loci within these regions does not differ from the total set of short SNPs although the small numbers of loci found within these regions may make any pattern difficult to observe. There are some patterns observed when looking at chromosome arms. Chromosome arm 3R, where *fru* is located, has 18 of the short intron loci. 14 of these were diagnosed as directional selection meaning the presence of balancing selection at the *fru* locus stands out.

Third and finally, heterozygosity could be maintained through linkage to other loci under selection, a phenomenon known as associative overdominance (AO; Ohta and Kimura 1970; Ohta 1971). AO occurs when each allele at a neutral locus is linked to a different recessive deleterious allele. Each focal allele and its associated deleterious allele will then act as a single haplotype and since the linked alleles are recessive, heterozygotes carrying chromosomes with different haplotypes will enjoy a fitness advantage as both deleterious recessive alleles are masked. This can create the illusion of balancing selection by heterozygote advantage at the focal locus (Ohta 1971). Differentiating between true balancing selection and AO is a major issue for experimental evolution studies, especially where the length of time is not sufficient to break down tight linkage relationships between loci (Schötterer et al. 2015, Franssen et al. 2015). AO has even been found to contribute significantly to the maintenance of polymorphisms in some studies (Becher et al. 2020; Gilbert et al. 2020; Schou et al. 2017). Furthermore, the risk of AO occurring is particularly high in this experiment, since the populations were created from a small set of inbred lines, leading to higher levels of LD and linkage over longer ranges. It is therefore possible that the balancing selection detected at *fru* or the short intron SNPs is not due to selection, but rather a result of AO.

The problem of accounting for apparent patterns of selection at short intron SNPs is not unique to this study as previous studies have also

struggled to find any difference between the behaviour of candidate selection loci and short intron SNPs (Tobler et al. 2013). That neutrality is observed in some cases is a positive result, especially since some of these loci may not be 'truly' neutral (Clemente and Vogl 2012). The suggestion of neutrality of short introns comes from between species comparisons where the low mutation rate in introns (Nolte and Schlötterer 2008) and their high AT content mean that short introns behave neutrally over long time frames (Parsch et al. 2010). However, it is not possible at this time to rule out that the patterns of balancing selection observed for the three proxy SNPs are due to AO. These potential confounding effects pose a problem for the results of this study and will be need to be accounted for in future to have complete confidence in the action of balancing selection at *fru*.

As the experiment progressed, recombination will have continued to break down the effect of AO likely resulting in more short intron loci behaving neutrally. This could be investigated if further sequencing can be performed. A larger data set of short intron loci could also be helpful, although identifying independent short intron loci which are also polymorphic in our DGRP lines has proved difficult. Alternatively, a model specifically simulating the AO would be useful to discern between this and balancing selection and I am currently working on developing such a model. For the time being, while the results of the three proxy SNPs clearly demonstrate a case for balancing selection, the fact that 44/100 putatively neutral short intron SNPs also appear to be under balancing selection means that the results pertaining to *fru* should be taken with some caution until this discrepancy is accounted for.

3.6.3 Evaluation of methodology

An important step of this project was the planning and performing of the pool sequencing. All analysis is based on pool-seq data, with simulations designed to mirror population sampling and sequencing coverage. If the pool-seq design is biased, subsequent conclusions are tainted (Schlötterer et al. 2014). To ensure a successful pool-seq experiment there were three main

things to consider: 1) the number of samples (pools) to be sequenced; 2) the number of individuals per pool; and 3) the sequencing depth (coverage) in each pool. The simulations performed here showed that the highest accuracy and lowest error in estimating the allele frequency occurred when using nine time points and keeping the sequencing effort constant (Figure 3.2). Combining this with the advice of Schlötterer et al. (2014) resulted in 48 flies per pool and average sequencing coverage of 40. In reality sequencing coverage for 2/3 of the proxy SNPs was lower than 40 (1181=21.3 and 1208=24.1). This is unfortunate, but expected due to the poor mapping of reads in regions near substantial indels. However, other studies have achieved high predicted accuracy using coverage values nearer to 20 (Bergland et al. 2014; Machondo et al. 2016, 2021). The level of coverage of the proxy SNPs is still adequate for results to be considered accurate and reliable. Based on a balance of the simulations performed, advice from both reviews and other published studies, and personal communication with experienced pool-seq users (Dmitri Petrov, Stanford University), I am confident that the sequencing approach taken has been effective.

Previous efforts to identify selection such as E&R studies have typically used traditional statistics like Tajima's D (Tajima 1989) and F_{ST} or used a comparison measure like a Cochran-Mantel test (Cochran 1954; Mantel and Haenszel 1959; Kofler et al. 2011, 2014). However, I have taken a different approach of using ABC to discern to diagnose the mode of selection occurring. This has provided a useful and efficient way of capturing the variation seen in the allele trajectories (Figure 3.4) and contrasting these with simulated data. The advantage of this method is that the summary statistics can be tailored to the experiment itself rather than relying on the assumptions of other statistics. Calculating one value for all 10 cages also takes advantage of the multiple parallel populations as any difference in measure necessitates a consistent difference shared by populations and is less prone to random fluctuations of individual cages (Graves et al. 2017).

The summary statistics also provide a straightforward method to compare sequencing and simulated data.

The development of the ABC approach in terms of selecting appropriate summary statistics was important part of this project (see Appendix C for details). It was essential to understand how the summary statistics differed between the selection models (Figures C.2 and C.3) and determine if they were capturing the variation they were designed to. Without this process the number and design of the summary statistics would have been very different and far less useful in differentiating modes of selection. Comparing the summary statistics values to the parameters (Figure C.1) was also a useful step in refining the models. One potential issue with the summary statistics is that they were initially designed with the original experimental design in mind and thereby make use of the difference in starting *fru* allele frequencies (0.9 or 0.1). While this works well for the *fru* locus, this could be part of the problem when investigating the short intron SNPs, many of which do not have such drastic differences in starting frequencies. However, in Chapter 4 I show that there is no relationship between the starting frequency, or difference in frequencies and their probability of balancing selection. It would still be useful to review the summary statistics to ensure they capture the variation expected for a range of experimental set ups to see if improvements could be made. The summary statistics could be further improved to form a general method for identifying the mode of selection acting at sites in the genome, similar to that demonstrated by Wittman et al. (2017) for identifying selection in seasonally fluctuating populations. By adapting simulations, summary statistics could be monitored to see how they respond to changes in selection, calculate which statistics are capturing the most useful information and develop further statistics.

The development of useful summary statistics was essential for the application of ABC. The results from model fitting cross-validation, Figure 3.4

show that the models can be differentiated from each other based on these statistics alone. I am therefore confident both in the ability of the selection summary statistics to capture the differences between the selection models, and in the result of the model selection process that *positive balancing selection* is the most probable selection model generating the allele frequency trajectories. ABC can be applied using a number of different algorithms. When a large number of summary statistics are used for ABC it has been advised to use a machine learning algorithm like neural net rather than methods like rejection or multinomial logistic regression (Blum and François 2010; Csilléry et al. 2010). This is due to difficulties sometimes encountered in the use of correlations and confusion between large numbers of summary statistics (Beaumont 2019; Csilléry et al. 2010). However, I found that the model fitting cross-validation did not improve with a neural net method. The multinomial logistic regression method employed was therefore used for model choice of the proxy *fru* SNPs since it is just as good and far less computationally intensive.

One issue with the methodology is that balancing selection was modelled as overdominance. While in Chapter 2 I state that the mechanism acting at *fru* is AP rather than overdominance, modelling the locus in this manner was a more straightforward way to model balancing selection since this uses only two parameters, *s* and *h*. Using the same pair of parameters as the other selection models enables straightforward direct comparison using ABC which would be difficult with a more complicated model of balancing selection such as AP. The estimated parameters support the conclusion that *fru* is under balancing selection, as the mean estimated parameter value of *h* is greater than 1 for all three SNPs (Figure 3.6) as would be expected under overdominance. Also, by simulating balancing selection as overdominance, I could use the estimated parameter values to calculate the predicted equilibrium frequency of the S allele (Figure 3.9). It appears that, assuming that selection is sufficiently strong compared to drift, polymorphism at these loci and the *fru* indel could be maintained for

prolonged periods of time, a hallmark of balancing selection. However, building and testing a model of AP using the fitness values from Chapter 2 would be a valuable continuation of this work

3.6.4 Future applications

I have discussed above potential improvements that could be made or further work relating to differentiating between AO and true balancing selection, and for investigation and refinement of the summary statistics. There are other potential applications of the data and methodology developed here which could allow further questions on the maintenance of genetic variation to be investigated.

Sequencing was not restricted to these sites alone, but was generated for the whole genome. The methodology performed here could therefore be expanded to detect balancing selection at other putatively antagonistic polymorphisms. Ruzicka et al. (2019) generated a list of 2372 putative sexually antagonistic from a GWAS of *D. melanogaster* lines. SA loci have also been proposed as a means of generating balancing selection and maintaining genetic variation for fitness (Pennell and Morrow 2013; Schenkel et al. 2018). The mode of selection acting on the candidates of Ruzicka et al. (2019) in my populations could be investigated in a similar way as for the short intron loci. Although not all candidate SA loci will give rise to balancing selection, a locus that is both identified as SA by the GWAS of Ruzicka et al., and is diagnosed as evolving under balancing selection in our populations, is more likely to represent a true SA locus and not be an artifact of correlation in the GWAS. Evidence of balancing selection at SA candidates would add to their case as true SA loci without the need to perform time-consuming phenotypic assays.

In Chapter 2, I hypothesised that *fru's* role as a transcription factor may influence selection at multiple sites in the genome through the regulatory cascade (Vernes 2014). The proposed differential expression of

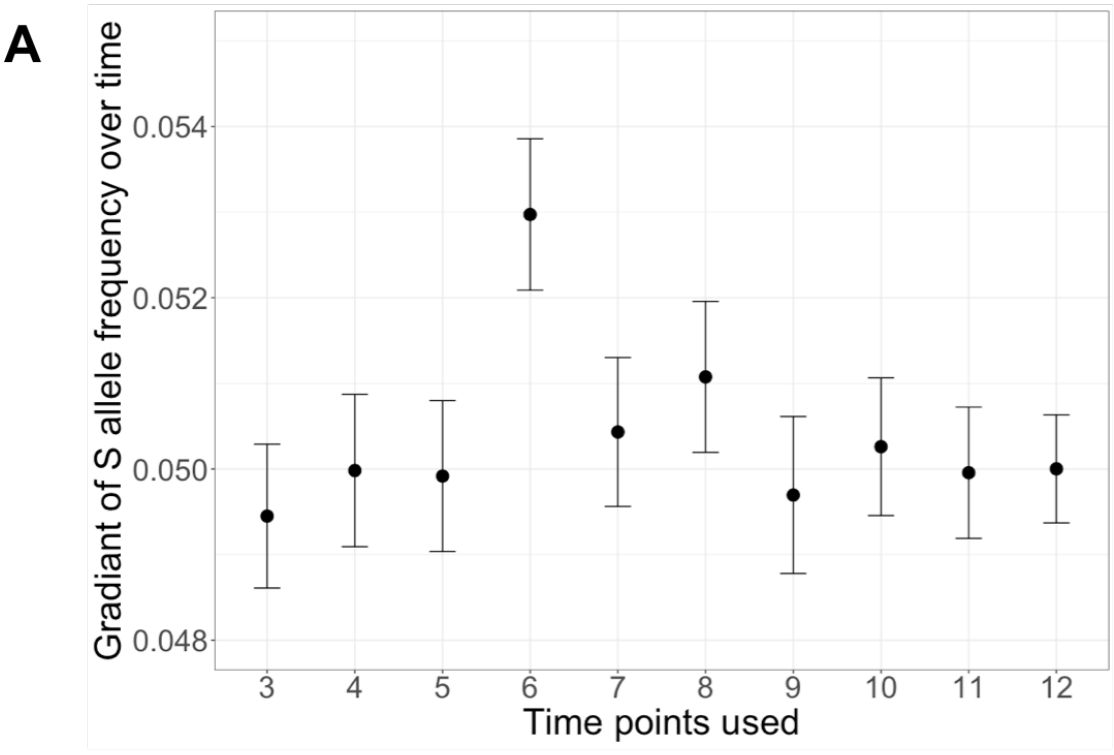
the three *fru* iso-forms expected from the indel polymorphism has the potential to impact on hundreds of sites across the genome. Therefore, any selection on the polymorphism could result in balancing selection being exerted over a much wider range than a single locus. To investigate this, the mode of selection acting at other loci within the *fru* regulatory cascade could be tested to see if these are affected by selection in a similar manner. This would add both support to the hypothesis in Chapter 2 and also demonstrate the power of antagonistic loci in maintaining genetic variation for fitness.

3.6.5 Conclusion

I have shown that the *fru* indel polymorphism is likely to be under balancing selection. This comes from the time series sampling and sequencing of flies from 10 parallel populations, and then a comparison of the resulting allele trajectories with simulations of the *fru* polymorphism under different selection models. I propose that the balancing selection observed at *fru* is a result of the AP detected in Chapter 2, meaning that AP is capable of maintaining genetic variation for fitness in populations. Yet this result is somewhat confounded by the substantial number of short intron loci which also displayed signs of balancing selection. Until the issue of balancing selection at these loci is satisfactorily resolved the conclusion that the *fru* polymorphism is maintained by balancing selection should be treated with some caution and this will need to be investigated in the future. The methodology described here could be developed to investigate the behaviour of loci within the *fru* regulatory cascade to examine the reach of balancing selection and also to investigate putatively antagonistic sites detected from association studies. Overall, this study indicates the possible occurrence of balancing selection in maintaining genetic variation in *D. melanogaster,* thus providing a unique case of active balancing selection occurring in real time, where both the fitness effects of the polymorphism are understood and where a historical signature of long-term balancing selection is detected in the genome.

3.7 Figures and Tables

Figure 3.1. A cage used to house an experimental population. The left-hand picture shows the cage from the side with the entrance on the left. The right-hand picture is taken from above showing the 12 bottles of food that each cage contained.


A
Gradiant of S allele frequency over time
0.054
0.052
0.050
0.048
3
4
5
6
7
8
9
10
11
12
Time points used

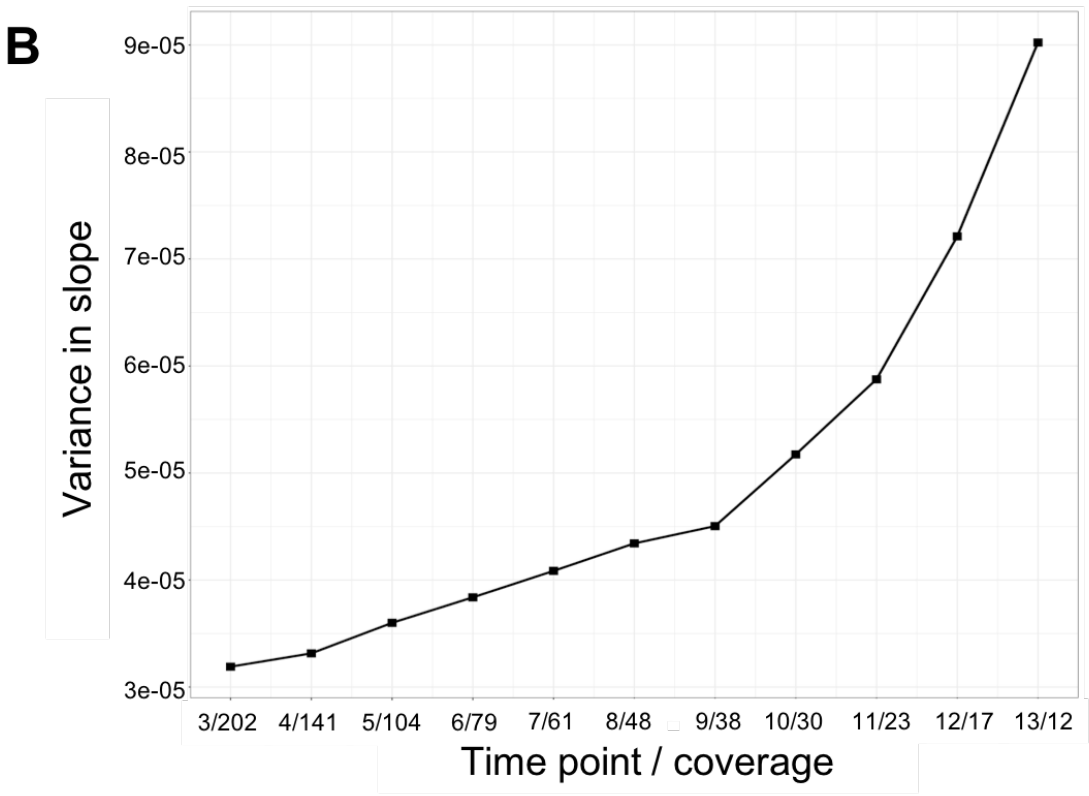

B
Variance in slope
9e-05
8e-05
7e-05
6e-05
5e-05
4e-05
3e-05
3/202
4/141
5/104
6/79
7/61
8/48
9/38
10/30
11/23
12/17
13/12
Time point / coverage

Figure 3.2. Results from simulations of variable pool-seq design. **A)** The gradient of the increase in the S allele measured by different numbers of timepoints. The true gradient is 0.05. Although fewer timepoints are nearly as accurate as more timepoints, there is more movement around the true value as the precise choices of the timepoints affect the measured gradient. **B)** Variance in slope against the number of timepoints and coverage depth used to estimate the gradient. The x-axis is shown as: number of timepoints/average sequencing coverage. The volume of sequencing data for each point, so those with more timepoints also have lower average coverage. Variance increases slowly with less coverage. There is a distinct change after nine timepoints at 38X coverage, as any further decrease in coverage results in poor estimate of allele frequencies.

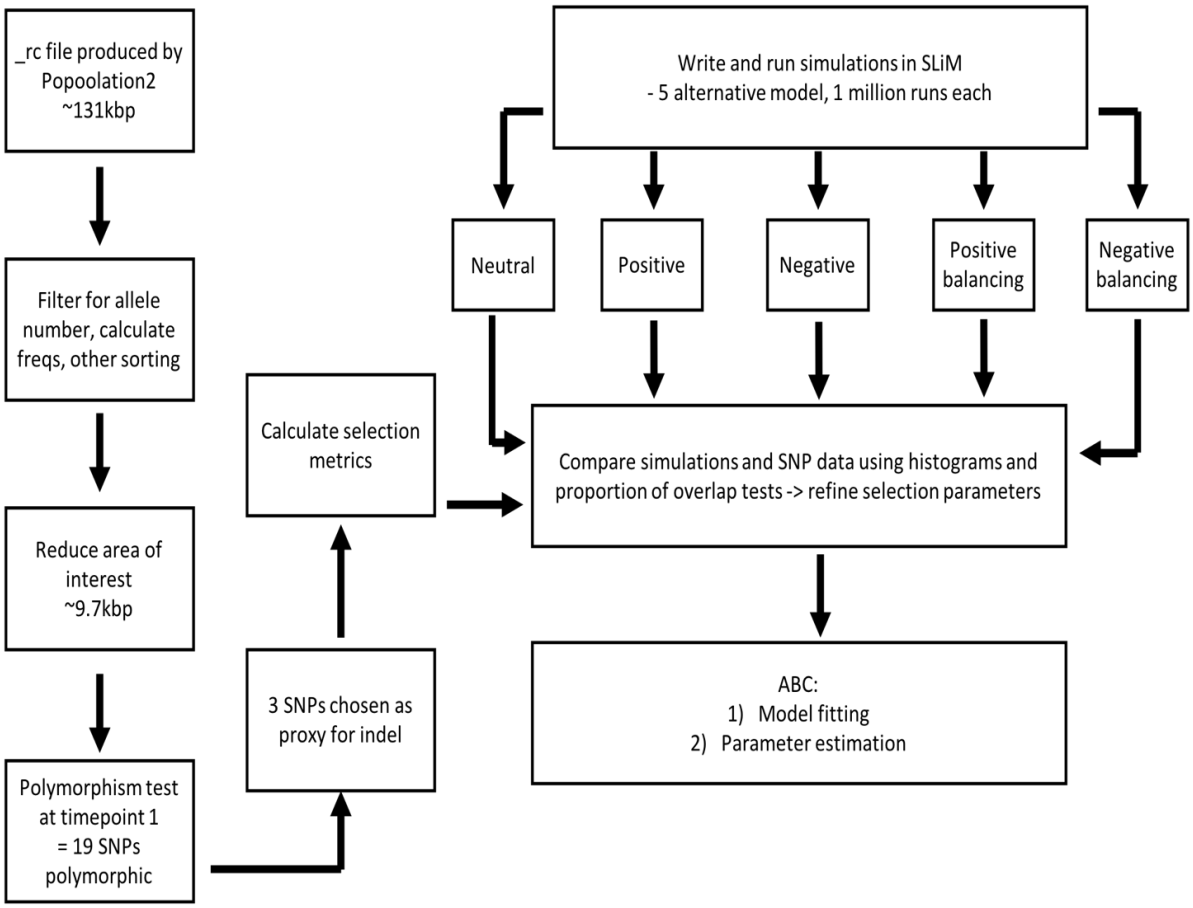


Figure 3.3. Flow diagram of steps taken to diagnose selection at *fru.*

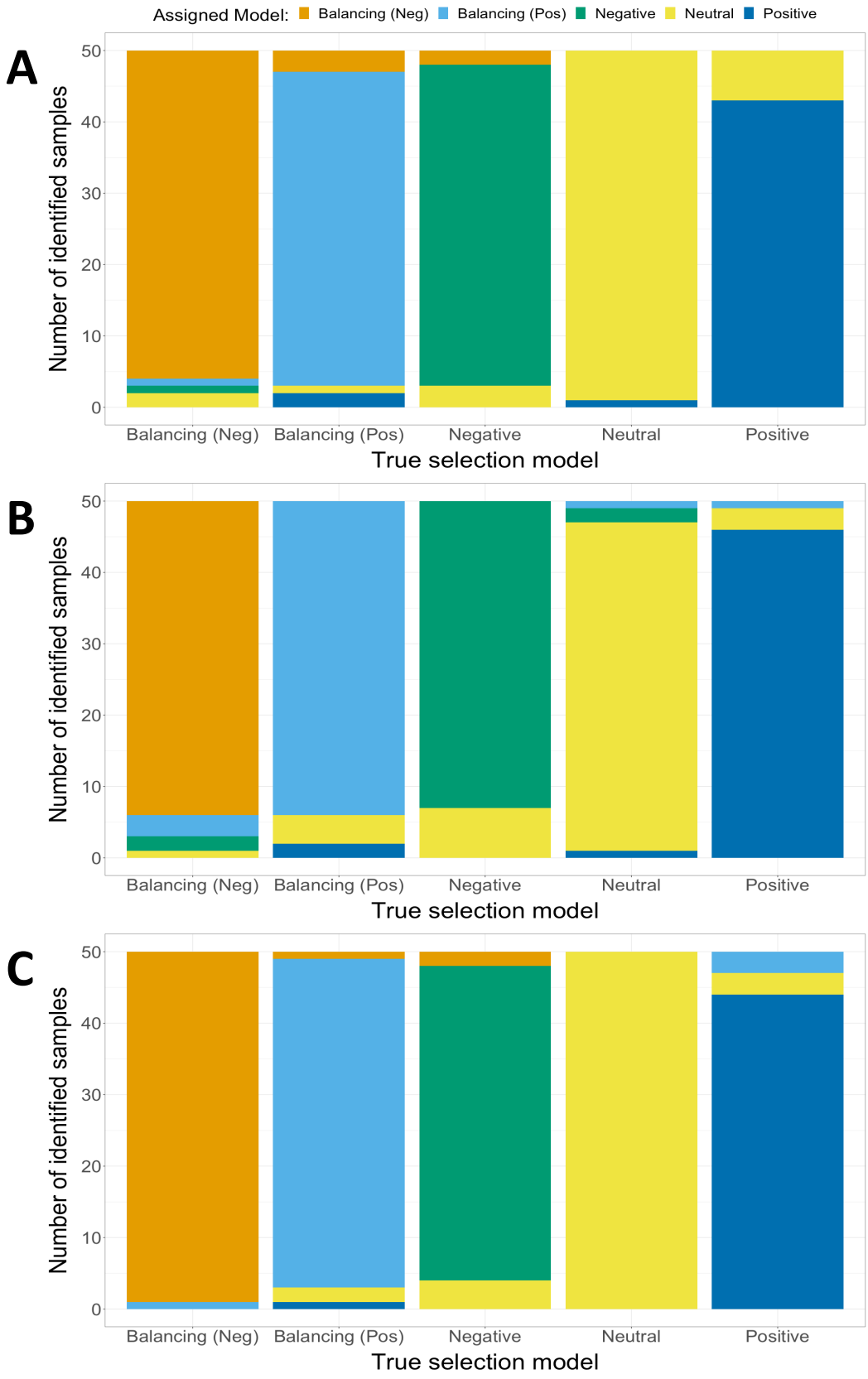

Assigned Model: Balancing (Neg) Balancing (Pos) Negative Neutral Positive
A
B
C
Number of identified samples
0
10
20
30
40
50
Balancing (Neg)
Balancing (Pos)
Negative
Neutral
Positive
True selection model

Figure 3.4. Model choice cross-validation confusion matrices. Each sample's true selection model identity by the total number of samples for each of the three SNPs (**A)** SNP 1181, **B)** SNP 1208, **C**) SNP 1521). The model class to which they were assigned by the cross-validation process is shown by colour. 50 samples were used per model. For every model and SNP, the majority of samples were assigned to their true model classification indicating that the models can be differentiated from each other based on their metric values alone.

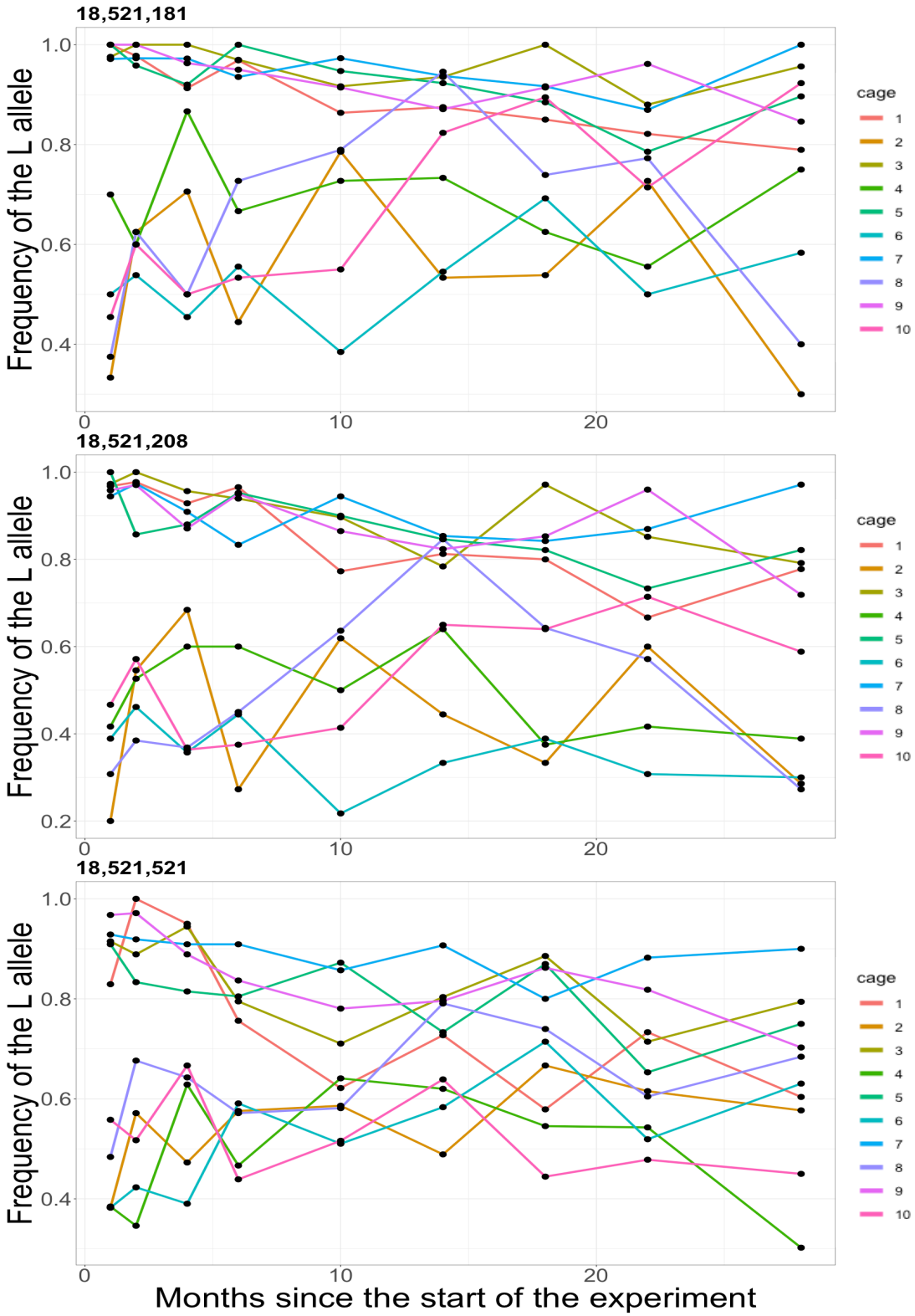

18,521,181
18,521,208
18,521,521
Frequency of the L allele
Months since the start of the experiment
cage
1
2
3
4
5
6
7
8
9
10

Figure 3.5. Allele frequency trajectories for the three proxy SNPs over the number of months during the experiment. SNPs are colour coded by cage number. Odd numbered cages started with a S allele frequency of 0.9 and even numbered cages with a frequency of 0.1. The location of each SNP appears above its frequency trajectory.

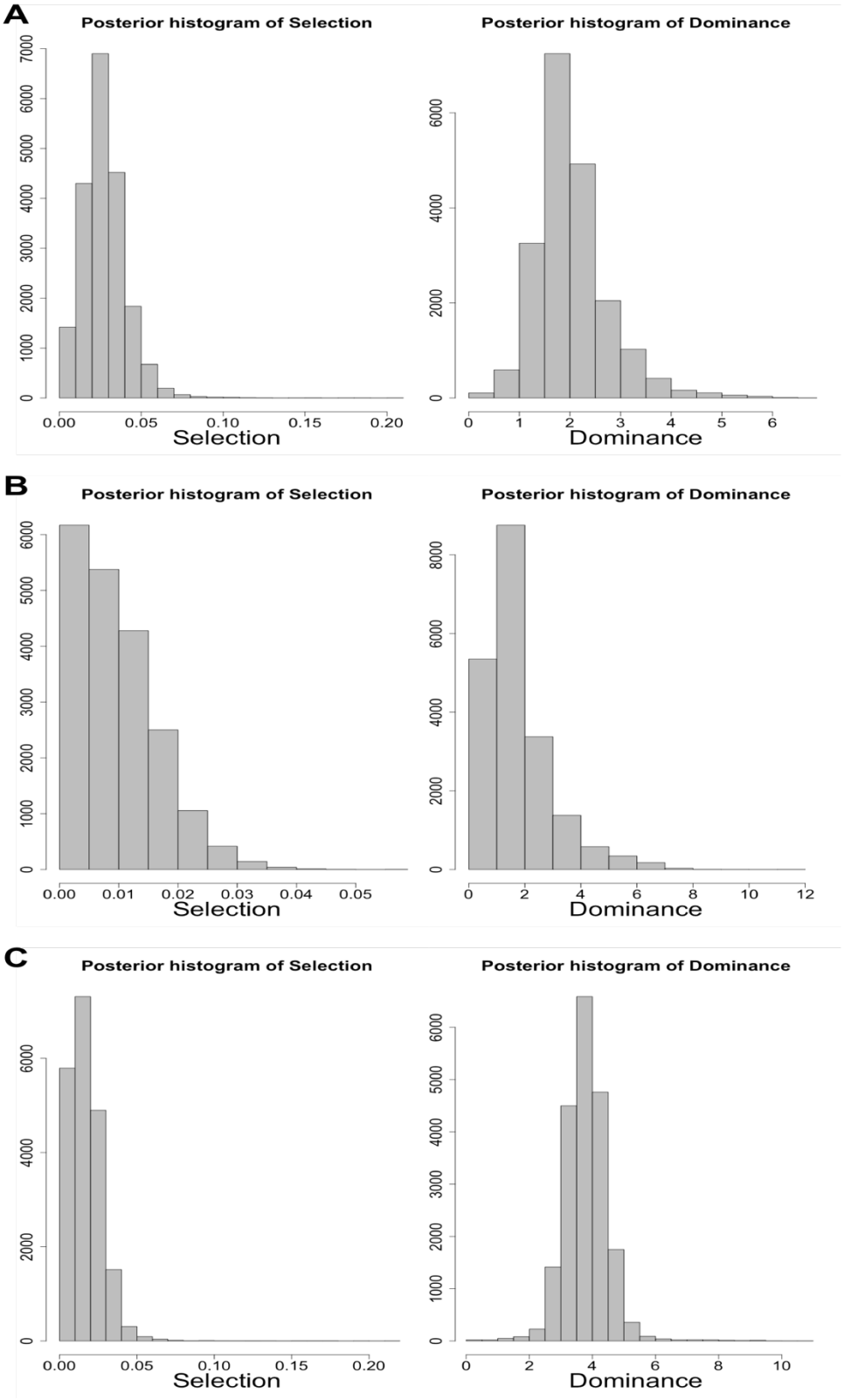

A
Posterior histogram of Selection
Selection
Posterior histogram of Dominance
Dominance
B
Posterior histogram of Selection
Selection
Posterior histogram of Dominance
Dominance
C
Posterior histogram of Selection
Selection
Posterior histogram of Dominance
Dominance

Figure 3.6. Histograms of the estimated parameter values. **A)** SNP 1181; **B)** SNP 1208; **C)** SNP 1521. All parameters were estimated using ABC and neural net method. Estimated values were all converted to be positive.

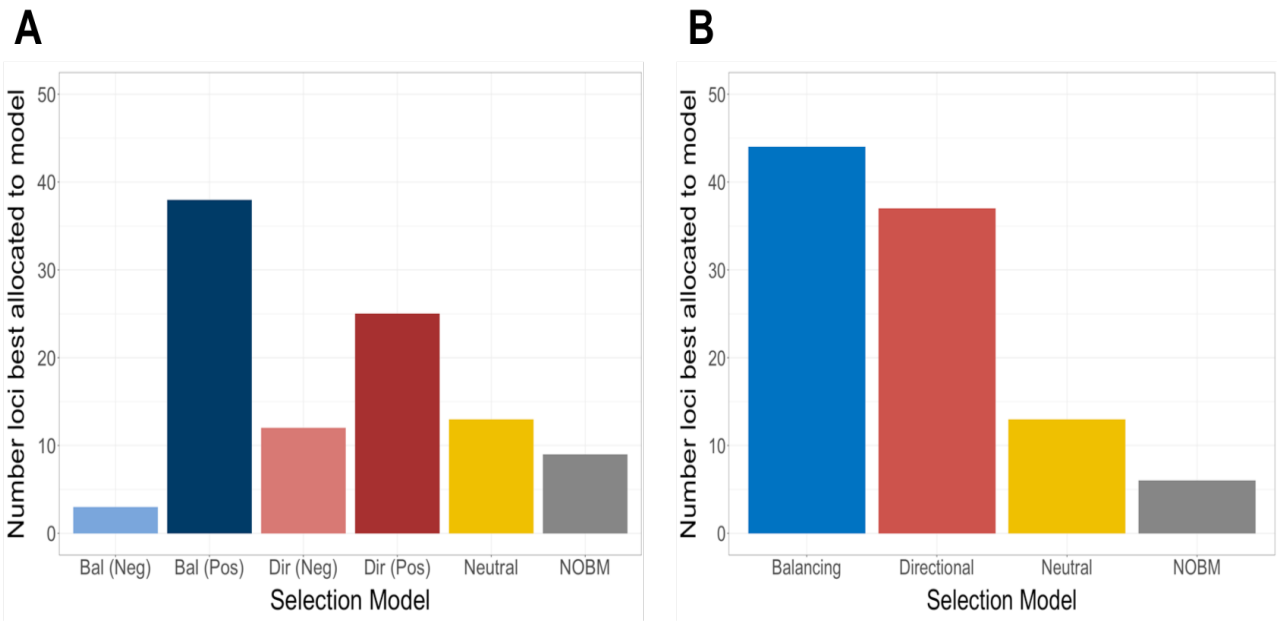


Figure 3.7. Number of short intron SNPs where the posterior probability is >0.5. **A)** Probability counts for all five selection models, plus no overall best model (NOBM). *Positive balancing selection* (Balancing (Pos)) has the highest number of loci fitting this model of selection. **B)** The two forms of balancing selection and the two forms of directional selection are combined to give four probability outcomes.

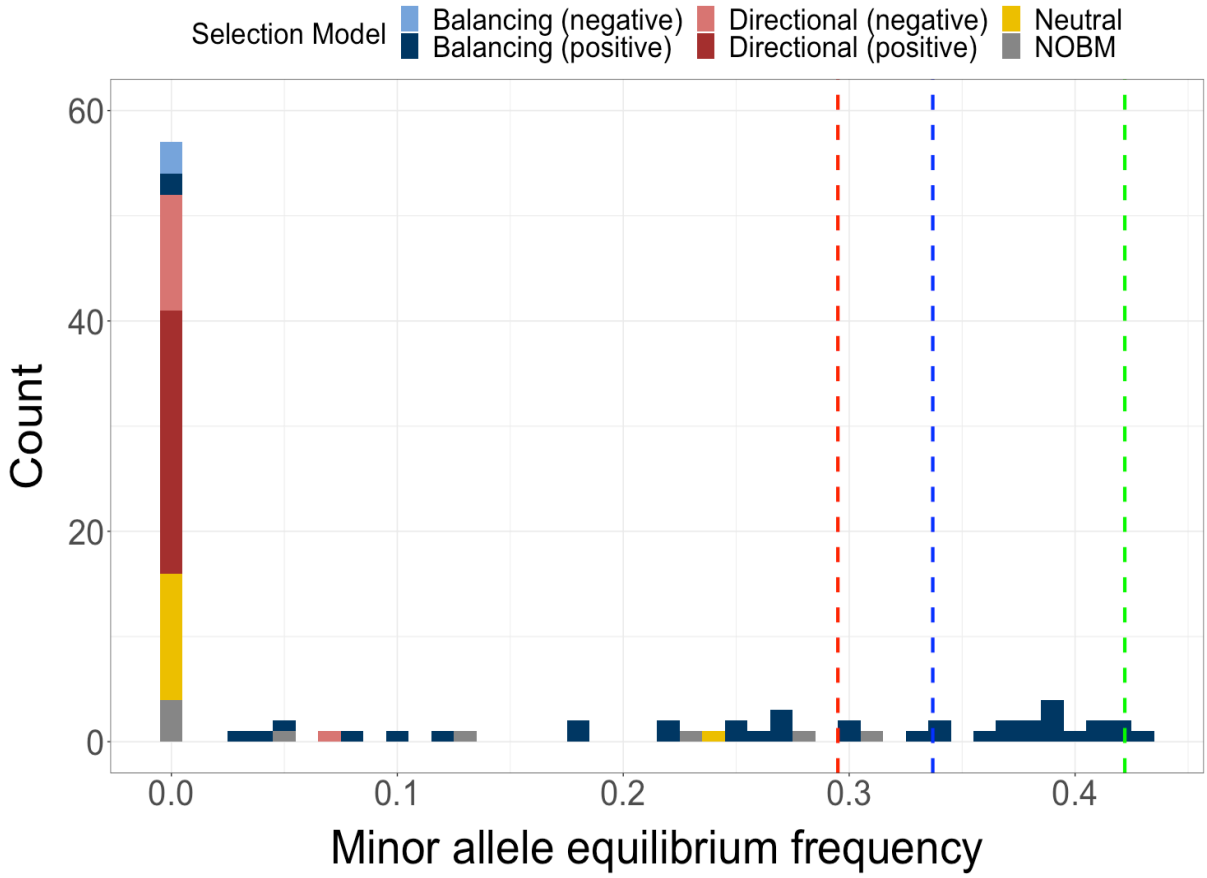


Figure 3.8. Predicted minor allele frequencies of short intron SNPs. Frequencies are calculated from the estimated parameter values. Loci are coloured by the selection model diagnosed by model choice posterior probability. The equilibrium frequencies of the three proxy SNPs are shown by the dashed lines (1181= blue, 1208=red, 151=green). While many short intron SNPs are predicted to become fixed, some have similar predicted equilibrium frequencies as the proxy *fru* SNPs.

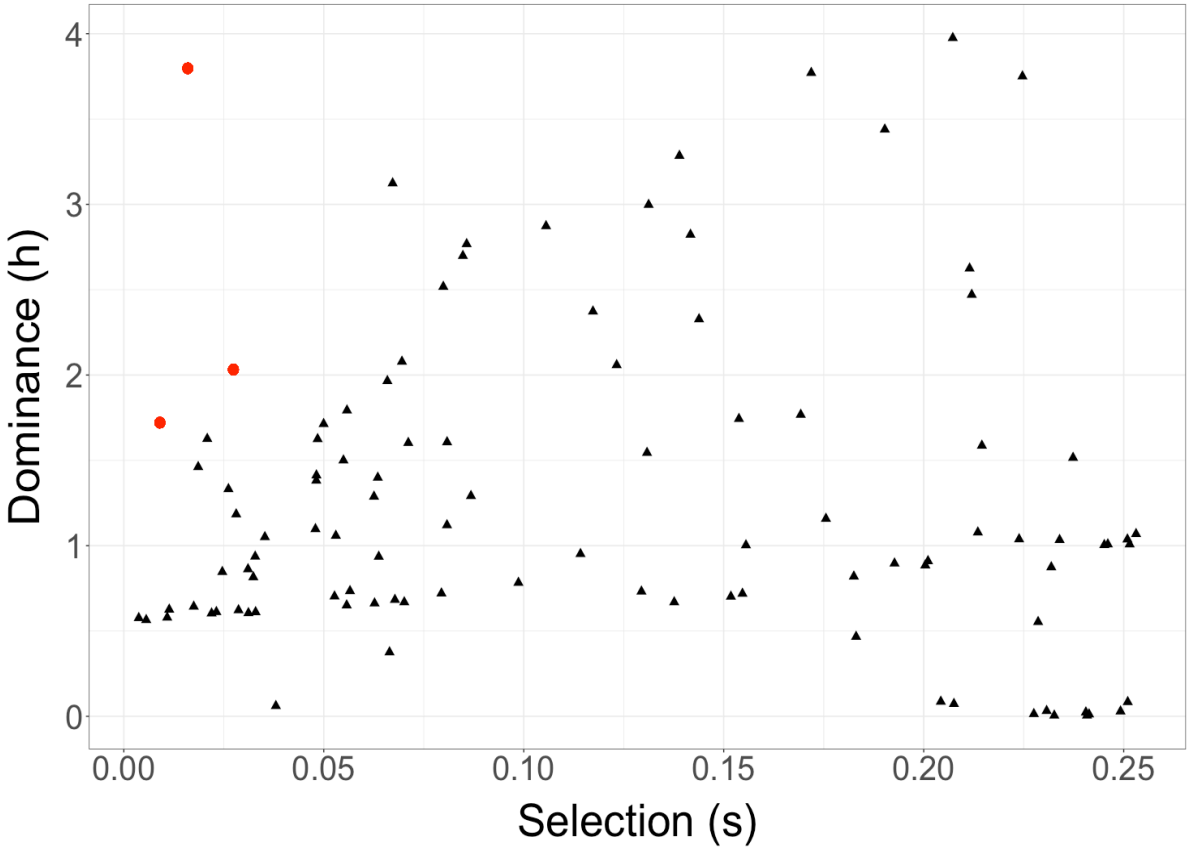


Figure 3.9. Selection (*s*) against dominance (*h*) for 100 short intron SNPs. While there is a large spread of values estimated for these loci, there are patterns also with clusters reporting a high *s* and *h*~0, or low *s* and intermediate *h*. The same measurements for the three proxy SNPs are shown by the red dots. These occur on the edge, away from the main distribution with a unique combination of low *s* and high *h*.

Lines fixed for S allele	Lines fixed for L allele
RAL-109	RAL-142
RAL-129	RAL-26
RAL-228	RAL-315
RAL-321	RAL-320
RAL-385	RAL-380
RAL-399	RAL-40
RAL-439	RAL-441
RAL-461	RAL-492
RAL-584	RAL-509
RAL-799	RAL-517
RAL-801	RAL-589
RAL-88	RAL-596
RAL-908	RAL-832
RAL-93	RAL-850

Table 3.1. Names of the 14 DGRP lines fixed for the S or L allele. These populations were then combined to establish the 10 cage populations using different ratios of the two fixed populations. Each cage population contains a mix of all the above DGRP lines.

Selection model	Selection coefficient, *s*	Dominance coefficient, *h*
Neutrality	0	0
Positive directional	0 – 0.25	0 – 1
Negative directional	-0.25 – 0	0 – 1
Positive balancing	0 – 0.25	1 – 5
Negative balancing	-0.2 – 0	-4 – 0

Table 3.2. Ranges of parameter values for the various selection models. Values refer to the focal SNP tracked by SLiM. The specific value used in each simulation is taken in a random uniformly distributed manner with the range.

SNP	Metric Number												
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
1181	0	56	0.278	-7.67E-05	0.076	0.3	0.058	-0.026	1.3	0.003	0.2	0.745	0.306
1208	0	56	0.190	-0.001	0.082	-0.3	0.068	-0.042	1.1	0.002	0.1	0.592	0.449
1521	0	56	0.289	-0.001	0.051	-0.3	0.029	-0.036	2.3	0.001	0.5	0.639	0.221

Table 3.3. Selection summary statistic values for the three proxy SNPs. Summary statistics are referred to by the number given during the descriptions. SNPs are labelled with the final 4 numbers of their position. Values are reported to 3 decimal places.

SNP	Model Type				
	Neutral	Positive	Negative	Balancing (positive)	Balancing (negative)
1181	0.0046	0.1795	<0.0001	**0.8147**	0.0012
1208	0.0127	0.0025	0.001	**0.569**	0.4157
1521	<0.0001	0.0004	<0.0001	**0.9977**	0.0019

Table 3.4. Posterior model probabilities of ABC model choice. Values >0.5 appear in bold and indicate the most probable model to have produced the observed data. Values are reported to 4 decimal places.

	Model Type									
	Neutral		Positive		Negative		Balancing (positive)		Balancing (negative)	
SNP	Distance	p-value	Distance	p-value	Distance	p-value	Distance	p-value	Distance	p-value
1181	16.112	<0.001	5.092	0.170	7.385	0.080	8.301	0.130	10.060	0.040
1208	12.214	<0.001	5.987	0.030	7.118	0.060	10.401	0.070	10.847	0.050
1521	20.193	<0.001	5.827	0.120	6.511	0.180	10.322	0.110	10.993	0.100

Table 3.5. Model goodness-of-fit for each SNP and selection model. Distance is a test statistic comparing the true SNP value to the model's posterior distribution. A p-value >0.05 indicates that the model fit is good. Values are reported to 3 decimal places.

	No. of loci with >0.5 model probability
Neutral	13
Positive	25
Negative	12
Balancing (positive)	38
Balancing (negative)	3
No best model	9
Neutral	13
Directional (total)	37
Balancing (total)	44
No best model	6

Table 3.6. The number of short intron SNPs diagnosed as under each mode of selection. Loci were assigned to a selection model if the posterior probability calculated by ABC was >0.5. If no model's probability was >0.5, no form of selection could be diagnosed and such loci are counted in the 'no best model' category. The above table is split into two, the top half treats the five selection models separately, while the bottom half combines the probabilities of the two directional and two balancing selection models.

Chapter 4

4. Investigating the occurrence of balancing selection at candidate SA SNPs

4.1 Declaration

Much of the data used in this chapter was produced by the same experiment described in Chapter 3, including section 3.4.2 ‘Cage population establishment and maintenance’ which was performed by Filip Ruzicka with help from Harvindar Pawar and Olivia Donaldson. The SLiM scripts for selection modelling were written in collaboration with Carl Macintosh. All other work presented here is my own.

4.2 Abstract

Sexual conflict emerges as the two sexes have different paths to achieving higher life-time reproductive success but also share most of the same genetic architecture. Sexual antagonism is a type of conflict occurring at a single locus, where the allele associated with greater fitness is different in each sex. This can result in balancing selection and maintain genetic diversity at such loci. Studies have predicted and found evidence for sexually antagonistic (SA) variation in many species. However, there are few clear examples where both the genetic and phenotypic variation of SA loci are understood, or where it is known that these loci are actively maintained by balancing selection. A recent study by Ruzicka et al. (2019) identified 2372 candidate SA SNPs which show signals of stable excess polymorphism across populations. In this chapter I experimentally test for the presence of balancing selection at these candidate SA SNPs using experimental evolution. Using a similar approach to Chapter 3, I track the frequency of 397 candidate SNPs over 56 generations to create allele trajectories for each. Comparing these trajectories to those from simulations of each SNP under different modes of selection allows the mode of selection most likely acting at each SNP to be diagnosed using Approximate Bayesian Computation. This found that >60% of SNPs appear to be under some form of balancing selection. I found no relationship between the SA effect sizes estimated by Ruzicka et al. (2019) and the probability of balancing selection in this study. There was a positive relationship between the predicted minor allele frequency of a SNP in both the cages and in two wild populations. I repeated the analysis looking at five SNPs which were chosen for editing with CRISPR for fitness assays. I found that two of these showed strong signs of balancing selection and are therefore prime candidates for investigation. This chapter shows that a majority of SA candidates are likely under balancing selection and thereby represent true SA SNPs. The analysis has several areas for future development which could improve the analysis performed.

4.3 Introduction

Sexual dimorphism, where individuals of two separate sexes display different morphological, physiological and behavioural phenotypes, is a widespread occurrence across many forms of life (Cox and Calsbeek 2009; Hedrick and Temeles 1989). This dimorphism is a phenotypic reflection of the divergence in the reproductive roles the sexes perform. This divergence, in turn, is ultimately a result of anisogamy, the sex difference in gamete sizes (Parker 1979; Trivers 1972). Males produce many, small, cheap gametes while females produce fewer, large, more expensive gametes. As a result, the path to maximising fitness is different for the two sexes (Parker 1972, 1979). For males, the limiting factor to high reproductive fitness is the number of matings they can attain, since their gametes are cheap to produce and they usually don't have any futher investment in any offspring. Females, on the other hand, have large and expensive gametes requiring substantial resources, meaning that they are typically more careful than males about the resources they invest in reproduction (Trivers 1972; Chapman et al. 2003). This was demonstrated by Bateman (1948) who showed that male fitness increased with the number of matings completed, while female fitness did not. The different paths to increased fitness have knock-on effects on any other trait that enhances the reproductive strategy of either sex, resulting in different phenotypic optima for traits in each sex (Connallon and Clark 2014a; Pennell and Morrow 2013). For example, it may be beneficial for males of a species to be large and colourful so that they can attract mates and fight off rivals, whereas for females, who do not need to do either of those things, it's better to be inconspicuous and smaller in order to avoid predation. Disagreement between the sexes of this kind in how to maximise fitness can result in different and conflicting selective pressures on each sex – a situation known as sexual conflict (Chapman et al. 2003).

Conflict between the sexes comes in two main forms: inter- and intra-locus sexual conflict (Pennell et al. 2016; Schenkel et al. 2018). Inter-locus

sexual conflict describes a situation where the sexes disagree about each other's optimum trait or strategy—and hence genotype at the loci underlying them (Chapman et al. 2003; Chapman 2006). For example, one trait in males may be selected to maximise fitness in a way that has detrimental effects on female fitness. Other traits in females may then be selected to resist such action leading to continued rounds of co-evolution or inter-sexual arms races (Van Valen 1973). A classic example of this dynamic is that of the seminal fluid peptides of *Drosophila melanogaster*. Males produce these to manipulate females into laying more eggs, boosting the fitness of the male while also resulting in shorter lifespan for the female (Chapman et al. 1995; Fowler and Partridge 1989). Experimental evolution in which female adaptation to male harm is prevented results in increased virulence of male mating, indicating that females are normally under selection to resist the effect of the seminal fluid proteins (Rice 1996).

Intra-locus conflict occurs when the sexes disagree over the optimum genotype at a single locus that underlies variation of a shared fitness trait (Bonduriansky and Chenoweth 2009; Pennell and Morrow 2013, Van Doorn 2009). Such loci are commonly referred to as sexually antagonistic (SA). To illustrate this, assume that we have a locus for a shared trait, such as body size, where the relationship between the trait and fitness is different in each sex, e.g. large males and small females have greater fitness. This locus has two alleles, A and B, and the phenotypic effect of each allele is the same in each sex. Let’s say that the A allele contributes to producing larger individuals than the B allele. This results in greater reproductive fitness for males that carry the A allele compared to B carrying males. The A allele is therefore selected for in this sex. However, the A allele carrying females have lower fitness than B carrying females. The B allele is thereby selected for in females, the opposite scenario compared to males. This means that the direction of selection for each allele is sex-specific, as neither is fittest in both sexes (Lande 1980; Kidwell et al. 1977). Since the locus is shared and has similar effects in both sexes, both sexes cannot simultaneously attain their

fitness optima, leading to some maladaptation in at least one sex and decreased average fitness for the population as a whole (Figure 4.1) (Cox and Calsbeek 2009; Bonduriansky and Chenoweth 2009). Furthermore, balancing selection may result (Kidwell et al. 1977) where neither allele becomes fixed and SA variation is maintained, contributing to the overall amount of genetic variation for fitness in the population (Rice 2000; Schenkel et al. 2018).

As the sexes mostly share the same genome, and have greatly different strategies to maximise fitness, there is potential for widespread SA genetic variation (Pennell and Morrow 2013). The balancing selection this may generate could be a major factor in maintaining genetic variation for fitness (Connallon and Clark 2014a; Kidwell et al. 1977; Rice 2000). As a consequence, SA variation may play a role in other phenomena, including speciation (Bonduriansky and Chenoweth 2009), gene regulation (Fisher 1931), ontogeny (Rice and Chippendale 2001), sex-chromosome formation (Vicoso and Charlesworth 2009), and genome structure (Charlesworth and Charlesworth 1980; Kirkpatrick 2016). SA variation also has been found in humans (Ruzicka et al. 2021) and may be particularly important for understanding sex differences relating to disease (Harper et al. 2021). Specifically, SA variation has been implicated in the maintenance of polycystic ovary syndrome (Casarini and Brigante 2014) and in the variable pathology of STIs (Wardlaw and Agrawal 2019). In these examples, balancing selection as a result of SA variation maintains alleles that are detrimental to the health and fitness of one sex because of positive effects they have in the other sex (Gilks et al. 2014; Van Doorn 2009).

A number of studies have demonstrated varying levels of evidence for SA genetic variation in a wide range of organisms, including mammals (Foerster et al. 2007; Mokkonen et al. 2011), birds (Merilä et al. 1997; Price and Burley 1994), reptiles (Svensson et al. 2009), insects (Chippendale et al. 2001; Fedorka and Mousseu 2004; Gibson et al. 2002; Rice 1992) and

dioecious plants (Delph et al. 2011). These studies are based on two main approaches. The first is to demonstrate that a specific trait has different and opposing effects on the fitness of each sex. Examples include testosterone production in bank voles (*Myodes glareolus*), where increased production is beneficial to males but harmful to females (Mokkonen et al. 2011), and bill colour in Zebra finches (*Taeniopygia gutta*), where males prefer females with duller bills and females prefer males with brighter bills (Price and Burley 1994).

Other studies use quantitative genetics (correlations between individuals that share variation) to show the presence of SA variation in general fitness measures such as lifetime breeding success. SA is assessed by measuring the intersexual genetic correlation for fitness (r_{MF}) which reflects the relationship between male and female fitness across genotypes (Connallon and Clark 2014a; Bonduriansky and Chenoweth 2009; Long and Rice 2007; Chippendale et al. 2001). When shared genetic variation produces similar effects on fitness in both sexes (sexually concordant selection), r_{MF} will be positive. Conversely, r_{MF} is negative when the shared genetic variation results in opposing fitness effects between the sexes, a sign of SA variation. For example, Chippendale et al. (2001) found a negative r_{MF} for sex-specific fitness among a sample of genotypes extracted from an outbred laboratory population of *D. melanogaster*, where genotypes that have high relative fitness in males tended to have low relative fitness in females and vice versa. Related approaches have demonstrated the presence of SA fitness variation in red deer (*Cervus elaphus*) (Foerster et al. 2007), and ground crickets (*Allonemobius socius*) (Fedorka and Mousseau 2004).

While the approaches above may detect evidence of SA genetic variation, they are not able to identify the individual genes and sequence polymorphism involved (Fry 2010; Pennell and Morrow 2013). This is a major limitation to understanding the role that SA plays in evolution and its contribution, through balancing selection, to maintaining genetic variation

(Bonduriansky and Chenoweth 2009). Two individual examples of SA genes have been identified through detailed studies of phenotypes and are well described: the ‘orange blotch’ phenotype caused by a mutation in the *Pax7* gene results in cryptic colouration in Lake Malawi cichlids, which is beneficial for females wanting to avoid predators, but detrimental for males who attract mates with bright colour patterns (Roberts et al. 2009; Seehausen et al. 2008); and the *Cyp6g1* gene of *D. melanogaster*. The allele DDT-R in this gene is SA in the absence of the pesticide DDT as it promotes female fitness but can have negative epistatic interaction on male reproductive success. When DDT is present the DDR-R allele is fittest for both sexes. (Daborn et al. 2002; McCart et al. 2005, Smith et al. 2011).

A solution to identifying SA loci on a larger scale and studying their general properties comes from the use of Genome-Wide Association Studies (GWAS). GWAS combine phenotypic measures (here of sex-specific fitness) and variant calls to identify loci associated with specific traits. GWAS are therefore a powerful method of identifying the genetic variation that contributes to fitness variation. A recent GWAS by Ruzicka et al. (2019) investigated SA in 202 hemiclonal lines from the laboratory adapted *D. melanogaster* population LH_M which has been maintained since 1996 and has been used extensively in investigation of SA fitness variation (Chippendale et al. 2001; Pischedda and Chippendale 2006; Rice et al. 2005). This produced the first list of candidate SA SNPs in *D. melanogaster*, identifying 2372 SNPs. Based on these candidates, Ruzicka et al. (2019) carried out additional tests to assess evidence for the signatures of balancing selection that one could expect for SA sites. Specifically, they compared the strength of SA effects with signatures of polymorphism in non-LH_M populations (to avoid confounding effects of increased GWAS power at more polymorphic sites). These revealed a positive relationship between the SA effect size and the minor allele frequency (MAF) (a measure of polymorphism) of the SNP in *D. melanogaster* populations in North Carolina (the DGRP, Mackay et al. 2012) and Zambia (DPGP3, Lack et al. 2015, both

population samples are whole-genome sequenced collections of wild-derived inbred lines). They also found elevated Tajima's D (another measure of polymorphisms) in SA regions compared to non-SA sites that matched the candidates in their frequency in the study population LH_M. These results show that the candidate SNPs may also be sites under balancing selection, which both increases their SA credentials and means that they may contribute to maintaining genetic diversity in *D. melanogaster*.

While the results of Ruzicka et al. (2019) have allowed us to get a first genome-wide appreciation of SA in flies, both the identification of the SA candidates in the first place and then the further population genomic comparisons of SA effect and polymorphism in other populations, are currently based on correlative associations. This is problematic because, even though care was taken to remove any confounding effects at more polymorphic sites, these efforts might not have been entirely successful. Much stronger proof would be provided if these candidate loci could be shown experimentally to be under balancing selection. Experimental evolution provides an approach to accomplish this (Schötterer et a. 2015). By tracking the frequency of SA candidate SNPs in populations over time, we can observe if their patterns of change in allele frequency are consistent with that expected under balancing selection. Balanced polymorphisms are expected to remain polymorphic over time and also to converge towards a stable equilibrium frequency. By comparing the allele frequency trajectories of the candidate SNPs to simulations of the SNPs under different selection regimes, as previously demonstrated in Chapter 3, it is possible to diagnose the mode of selection that the SNPs experience. The diagnosis of balancing selection will add support for the notion that the SNPs are true SA loci and evolve under balancing selection. Such an approach combining GWAS and experimental evolution has been used previously to investigate the genes responsible for resistance to *Drosophila* C virus (Magwire et al. 2012; Martins et al. 2014), and to identify the loci behind variation in courtship song in *D. melanogaster* (Turner and Miller 2012; Turner et al. 2013).

In this chapter, I use an experimental evolution approach to assess selection at 397 of the SA candidate SNPs identified by Ruzicka et al. (2019). My analysis makes use of the sequencing data generated from the cage populations in Chapter 3, and I employ the methodology developed there to diagnose the mode of selection acting at these SNPs. I find that >60% of the candidate SNPs are diagnosed as being under balancing selection. I then relate the findings of selection at SA SNPs to properties of these loci calculated from the GWAS of Ruzicka et al. (2019) and to two wild populations. This forms a test of balancing selection at these loci to see if patterns of selection are consistent and predictable. I also investigate selection at five SNPs chosen to create gene-edited lines using CRISPR to see which of these should be prioritised for fitness assays.

4.4 Methods

4.4.1 Experimental design, sequencing, and bioinformatics

Sequencing data were obtained from the same experimental populations of *D. melanogaster* as described in Chapter 3. All treatment of replicate populations, sampling, DNA extraction, and bioinformatic analysis (sections 3.4.1-3.4.4) were the same as for all data presented here.

4.4.2 Diagnosing the mode of selection at candidate SA loci

The SA GWAS by Ruzicka et al. (2019) carried out in our laboratory identified 2372 putatively SA SNPs. Popoolation2 (Kofler et al. 2011) was used to extract the polymorphic SNPs from a whole genome BAM file of the pool-seq data for the putative SA candidate loci identified by Ruzicka et al. (2019). This created a _rc file of allele counts for our 90 pools covering the SA SNPs with sequencing variation. These SNPs were cross referenced with those loci expected to be polymorphic based on the DGRP data. This produced a reduced list of 442 candidate SA loci which were both predicted to be polymorphic from the DGRP data and were found to be so in our cage populations. Loci with more than two allele states were removed. Further data manipulation was performed to first extract the read coverage depth at each locus for each sample and then to convert allele counts to frequencies. The same 13 selection summary statistics used in Chapter 3 were then calculated for these SA candidates

Starting frequencies of loci in the cage populations were estimated from the whole genome sequence data provided from the DGRP lines. This was obtained from http://dgrp2.gnets.ncsu.edu/data.html where there is a vcf file containing whole genome sequences of all 205 DGRP lines. I filtered the vcf file using vcftools (Danecek et al, 2011) to create two subset files, one containing only data for the 14 lines fixed for the S allele and another dataset with the 14 L allele lines. The lines included in each set are in Table 3.1. vcftools then calculated the allele frequency at each locus across the 14

lines. The frequencies of the two sets were then combined in two different ratios, lines fixed 9:1 for S:L and the other 1:9 S:L. Assuming that all lines in each set contribute equally during the round robin cross (see section 3.4.2 – Chapter 3) and were allocated to the cages equally, these calculated frequencies should be the starting frequencies of each SNP when the cages were established. The list of loci was trimmed to remove loci where either allele was fixed and therefore there was no expected variation in the cage populations. This was trimmed again to remove those loci with expected starting frequencies either <0.025 or >0.975. Such loci are close to fixation already and therefore may fix during the set-up of cage populations or in the first few generations of the experiment due to drift rather than any selective forces. The output of this should be a full list of all polymorphic loci and their expected starting frequencies for each set of cages.

Selection model simulations were performed using SLiM (Haller and Messer, 2019). Five models were performed for each SA SNP: a *neutral* model, *positive selection*, *negative selection*, *positive balancing selection*, and *negative balancing selection*. These model types were formed using different ranges of the selection parameters, the selection coefficient, *s*, and dominance, *h*. The range for parameters used in each model type is shown in Chapter 3 Table 3.2. For each simulation the values of *s* and *h* were drawn from a uniform distribution and the value of each was recorded. Simulations were performed for 56 generations, the estimated length of the experiment which ran from July 2017 to November 2019. Simulations contained 10 subpopulations with no migration between them to represent the 10 cage populations in the study. The effective population size of each subpopulation was 1000. For each SA SNP each model was initialised with a different starting frequency using the value estimated from the DGRP lines. The allele frequency was then recorded for each of the 10 subpopulations after the end of generations 2, 4, 8, 12, 20, 28, 36, 44 and 56 representing the sampling points of the sequenced flies. These were then saved and passed to R (R

Core Team, 2019) for analysis. One million simulations were run for each model type resulting in five million simulations per candidate SNP.

Allele frequencies were then treated to two rounds of sampling to represent (i) the sampling of 48 flies from each of the cage populations included in the sequencing, and (ii) the read coverage at each site, from which the allele frequencies from Popoolation2 are calculated. 13 selection summary statistics were calculated for each simulation to describe the allele frequency trajectories over time (details of these summary statistics are described in section 3.4.4.1 in Chapter 3). The R package 'abc' (Csillery et al., 2012) was used to perform model choice to find the selection model with the greatest probability of having produced the real data of each SNP. Probabilities of selection models of the same type of selection, balancing or directional, were combined, irrespective of their direction. If the probability was >0.5 for any particular model that type of selection was determined to be the model responsible for the allele frequency dynamics at that locus. If no model recorded a probability >0.5 the best selection model could not be determined. Parameter estimation was also performed using 'abc' to estimate the range of *s* and *h* values most likely to have generated the real data. These values were converted to positive values ($s' = -s/(1 + s)$ and $h' = 1 - h$) for an allele under positive balancing selection to make comparisons easier. A diagram depicting the stages used in the analysis is shown in Figure 4.2.

4.4.3 Comparing experimental and GWAS results

To see if the results of Ruzicka et al. had any predictive power for the probability of balancing selection in this study, I used linear modelling to test for a relationship between the SA effect size estimated from the GWAS and the total posterior probability of balancing selection at each locus. The SA effect size was converted to an absolute value and the total posterior probability of balancing selection was normalised using the 'orderNorm'

transformation technique from the R package ‘bestNormalize’ (Peterson 2021; Peterson and Cavanaugh 2020).

To test if signatures of balancing selection measured in this study were consistent with those found in other populations, I tested for a relationship between the total posterior probability of balancing selection and the values of regional Tajima’s D (Tajima 1989) in the DGRP (Mackay et al. 2012) and DPGP3 (Lack 2015, 2016) populations. These values were calculated by Ruzicka et al. (2019) over a region of 1000bp around each SA candidate SNP. Higher values of Tajima’s D are associated with greater levels of polymorphism and thereby regions more likely under the influence of balancing selection. As a second measure of the maintenance of polymorphism I also tested if there was a relationship between the equilibrium minor allele frequency (MAF) in the cage populations, as predicted from the estimated parameter values, and the MAF reported at that locus in the DGRP and DPGP3 populations. I also tested if there was a relationship between the predicted equilibrium MAF and the SA effect size of each locus to see if this was predictable from the GWAS results.

To ensure that the experimental set up or the ABC analysis did not bias the results, I checked for a relationship between the mean starting allele frequency of each locus and its predicted MAF. Finally, a relationship between the difference in starting frequencies of the two sets of cages and the predicted equilibrium MAF was investigated. Both the difference in starting frequencies and the mean starting frequency were normalised using the ‘orderNorm’ technique. All tests were performed using linear models which included the chromosome arm where the locus was located, either 2L or 3L as an additional factor.

4.4.4 CRISPR SNPs

Eight putative SA SNPs were selected from the dataset of Ruzicka et al. (2019) by Ted Morrow (Karlstad University) and Jon Harper (University of

Sussex) to create a series of genetically modified fly lines. These lines are identical apart from at the SNPs chosen which were edited using CRISPR/Cas9 (referred to simply as CRISPR) genome editing technology to contain alternative alleles of the candidate SA loci. CRISPR allows for precise alteration of the genome, even by a single nucleotide (Doench 2018) and is derived from a defence mechanism against phage found in many strains of bacteria and archaea, which recognise DNA sites called Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) (Marrafini and Sontheimer 2010). At such sites, the Cas9 protein induces a double strand break, killing the phage. This is exploited by using two Cas9 proteins, with different user-defined target sequences, to excise and replace a portion of DNA with any desired alternative sequence (Lamb et al. 2017; Xi et al. 2015). Therefore, the creation of tailor-made modified organisms can be achieved, enabling the study of the fitness impact of small genetic variations (Doench 2018). This is a much more precise method of introducing genetic material than using introgression to create hemiclones (Abbot and Morrow 2011; Rice et al. 2005) and means that any fitness difference observed is directly due to the introduced genetic material (Doench 2018). CRISPR was performed using a two-step process by WellGenetics (Taipei City, Taiwan). SNPs were chosen based on the strength of the antagonistic signal detected by Ruzicka et al. (2019), if they occurred within protein coding genes, represented non-synonymous mutations and to cover the four main chromosome arms and the X-chromosome. The loci chosen are shown in Table 4.1.

I investigated the behaviour of the eight SNPs chosen for the creation of CRISPR lines within the cage population experiment. The expected starting frequencies calculated from the DGRP lines above were examined, but only two out of the eight SNPs were covered by the DGRP sequencing data, both being polymorphic in our cages. However, the _rc file output from Popoolation2 contains sequencing information covering all 8 selected CRISPR SNPs at frequencies not indicative of sequencing error. To make use of the information from more than just the two loci covered by the DGRP

data I altered the approach used above, ignoring the DGRP starting frequency estimates. I instead focused only on the 54 generations covered by the sequencing data generated from Popoolation2 taking the first sampled timepoint as the starting point of the experiment. I removed those SNPs which had more than two allelic states, of which there were two: 2L:18,998,623 and 3L:6,210,720. I also restricted myself to analysis of autosomal SNPs, thereby removing SNP X:2,937,496. Five SNPs remained to be investigated. Read coverage data were extracted for those SNPs and read count data from Popoolation2 converted to allele frequencies. Functions to calculate the 13 selection summary statistics had to be adapted to fit the shortened time frame of the analysis. This required some measure of the allele's starting frequency. For this I took the mean allele frequency at the first sampled timepoint of each set of five cages, i.e. the mean of cages 1-5 and the mean of cages 6-10. These values were used as the starting frequencies for these cages and are presented in Table 4.2. Summary statistics were then calculated for the five SNPs.

Simulations were run for these five SNPs using SLiM 3 (Haller and Messer, 2019). As before, five selection models were performed: *neutrality*, *positive directional selection*, *negative directional selection*, *positive balancing selection*, *and negative directional selection*. Selection models were specified with different ranges of *s* and *h* as before, drawn from a uniform distribution. Simulations were run for 54 generations with a population size of 1000. 250,000 simulations were run for each CRISPR SNP. Selection summary statistics were then calculated for each simulation using the same adapted functions as for the real sequencing data. ABC was performed using the abc package in R (Csilléry et al. 2012) to diagnose the mode of selection acting at each SNP. Model choice was performed using a neural net algorithm and 5% tolerance, while parameter estimation also used the neural net approach and using both model choice and parameter estimation calculations. A diagram depicting the stages of this analysis is shown in Figure 4.3.

4.5 Results

4.5.1 Diagnosing the mode of selection at SA loci

A total of 397 candidate SA SNPs were investigated. The most common selection model diagnosed was balancing selection with a total of 251 (63.22%) loci being diagnosed as evolving under either positive or negative balancing selection (see Table 4.3 for details). For the remainder of the SNPs, 55 (13.85%) were diagnosed as showing either form of directional selection, and 52 (13.1%) were behaving neutrally. For the remaining 39 (9.82%) SNPs, no model reported a probability value greater than 0.5 (Figure 4.4 and Table 4.3). The SA candidate SNPs tested here were selected from two of the four major autosomal chromosome arms, 2L and 3L. There were only minor differences between the two arms in the proportion of SNPs diagnosed as being affected by each type of selection. This was the case when distinguishing between the major types of selection (balancing and directional), neutrality and undetermined (χ_3^2=1.207, p=0.751) but also when considering all individual selection models (χ_5^2=3.97, p=0.554). While both chromosome arms were similar in their patterns of selection, the proportional representations of selection types inferred for SA SNPs here were different from those detected in short intron SNPs in Chapter 3 (major selection types: χ_3^2=100.506, p<0.0001; individual selection models: χ_5^2=118.597, p<0.0001). 17 of the SA candidate SNPs are located within the span of the In(2L)t inversion (2L: 2,225,744 – 13,154,180). Nine of these loci were diagnosed as under balancing selection, three as directional, three were neutral, and for two no selection model had a posteriori probability >0.5.

Parameter estimates varied between the 397 candidate SA loci. The median posterior mean selection coefficient, *s*, was 0.104 and point estimates ranged between 1.04×10^{-4} and 0.25. Posterior mean dominance, *h,* had a median of 1.455 with values ranging between 0.029 and 4.279. There was no difference in the posterior mean *s* between SA loci on the two chromosome arms (2L: *s*=0.11, 3L: *s*=0.113; Welch's 2-sample t-test: t=-

0.345, df=176.36, p=0.73) or for *h* (2L: *h*=1.689, 3L: *h*=1.68; Welch's 2-sample t-test: t=0.077, df=167.93, p=0.939). Estimated parameter values were used to calculate the predicted equilibrium minor allele frequencies of each SNP, and are shown in Figure 4.5. This shows that the majority of SNPs that were diagnosed as exhibiting balancing selection also had estimated parameter values that predict that the SNP will not become fixed.

4.5.2 Comparing experimental and GWAS results

A number of comparisons were made to test for relationships between the GWAS data of Ruzicka et al. (2019) and the findings of this chapter. The aim was to uncover any ability the GWAS data had in predicting the behaviour of the SNPs as calculated by ABC and if selection inferred here was consistent with that in other populations. All analyses accounted for potential differences between chromosome arms, but since none were detected these results are not described here (Results from all models is presented in Table 4.4).

First I assessed the relationship between the size of the SA effect of each SNP calculated from the GWAS of Ruzicka et al. (2019) and the evidence for balancing selection from the ABC. SA loci can give rise to balancing selection when the effects on male and female fitness are equal. In such a case, SNPs with more strongly antagonistic fitness effects would show clearer evidence of balancing selection in short-term frequency change. Yet there was no relationship between the SA effect size from the GWAS and the posterior probability of balancing selection from the ABC ($F_{1,394}$=0.231, p=0.631) (Figure 4.6A). There was also no relationship between the SA effect size and the equilibrium minor allele frequency (MAF) calculated from estimated selection parameter values ($F_{1,394}$=0.024, p=0.877; Figure 4.7A).

Ruzicka et al. (2019) found that the SA effect size and Tajima's D calculated from their GWAS correlated with population genetic measures of balancing selection from natural populations of *D. melanogaster.* I also tested

for an association between the same measures of balancing selection from wild populations and evidence for balancing selection from the ABC directly. Specifically, I focussed on two measures, Tajima's D and the MAF, calculated in a North American population (DGRP) and a Zambian population (DPGP3) in the species' ancestral range. Linear models provided no support for a relationship between the probability of balancing selection and Tajima's D from DGRP ($F_{1,394}$=2.114, p=0.149) (Figure 4.6B). However, there was an indication of a possible relationship between the probability of balancing selection and Tajima's D calculated in the DPGP3 lines ($F_{1,394}$=3.659, p=0.056) (Figure 4.6C). Contrasting with these somewhat equivocal associations, I detected a positive relationship between the MAF predicted from selection parameters inferred in the ABC and MAF estimated in the two wild populations. This was true both in the DGRP ($F_{1,390}$=32.609, p<0.0001; Figure 4.7B) and in the DPGP3 lines ($F_{1,359}$=27.78, p<0.0001; Figure 4.7C). For both sets of lines, roughly 7% of the variance in the predicted cage MAF is explained by the MAF value from population genomic samples. The relationship between the estimated cage MAF and the MAF calculated for the two wild populations was maintained even after removing SNPs with low starting frequencies (see below) (DGRP: $F_{1,346}$=20.08, p<0.0001; DPGP3: $F_{1,321}$=15.462, p<0.001).

I ran a number of additional models to check that the above results were not biased by the experimental set-up and analysis. Especially, I wanted to rule out that polymorphism in wild population would translate into greater variability in my cage population, which in turn might make the detection of balancing selection more likely and affect the predicted equilibrium MAF. There was no relationship between the estimated starting frequency and the predicted MAF, neither across all SNPs ($F_{9,337}$=1.65, p=0.1; Figure 4.8A) nor across those SNPs where the predicted equilibrium MAF was greater than 0, i.e. where polymorphism was expected to be maintained ($F_{9,242}$=0.0001, p=0.99; Figure 4.8B). Initially there did appear to be a positive relationship between the starting frequency and the predicted

equilibrium MAF at that locus ($F_{1,390}$=22.19, p<0.001). This relationship is due to loci with starting allele frequencies less than 0.15 being fixed early on during the experiment. This leads to bias in the mode of selection that is diagnosed at that SNP towards non-balancing selection and a predicted equilibrium MAF of zero. Finally, there was no relationship between the difference in the starting frequency between the two sets of cages and the predicted equilibrium MAF ($F_{1,390}$=1.274, p=0.26).

4.5.3 CRISPR SNPs

In reporting these results, I refer to the five CRISPR loci using the associated gene name (see Table 4.1). Of the five loci chosen for the creation of CRISPR lines and investigated here, two loci, Ugt36E1 and CG8399, exhibited a high (>0.5) probability of being under balancing selection during the 54 generations of evolution (Ugt36E1: post. prob.=0.7673, CG8399: post. prob.=0.767). Locus CG10477 reported a model choice probability >0.5 for the neutral evolution model (post. prob.=0.8268), and therefore does not appear to be under selection in this experiment. For the final two CRISPR loci, Strn-Mlck and CG34133, none of the five selection models had greater than >0.5 probability of having produced the observed summary statistics. For Strn-Mlck, the closest models were *positive balancing selection* and *positive directional selection*, with posterior probabilities of 0.4867 and 0.4889 respectively. For CG34133, the models with the highest posterior probabilities were *neutrality* and *positive directional selection* with values of 0.4625 and 0.4348 respectively. Full model choice results for the five loci are shown in Table 4.5. Parameter estimation was performed for all five loci. The mean estimated values were consistent with the mode of selection diagnosed by the model choice step above. Full parameter estimation results are presented in Table 4.6.

4.6 Discussion

4.6.1 Diagnosing balancing selection

Ruzicka et al. (2019) produced the first association study that identified candidates of SA genetic variation in *D. melanogaster*. They also demonstrated that their SA candidates showed signatures of balancing selection in natural populations of *D. melanogaster*, including excess polymorphism (compared to frequency-matched non-SA sites) and elevated Tajima's D. To corroborate their population genomic analyses, and thus gather additional support for SA at these loci, I have tested for signals of balancing selection at a subset of their candidate loci, using the experimental cage populations of *D. melanogaster* described in the previous chapter. Sexual antagonism can lead to balancing selection as alternative alleles are favoured by each sex. Therefore, loci identified as SA candidates that also show signals of balancing selection in the experimental populations are more likely to be true SA loci and contribute to maintaining genetic variation. However, it should also be noted that candidate SNPs that do not show clear signs of balancing selection are not discounted as SA, but rather the detection of balancing selection enhances the case for those that do. The selection properties of candidate SA SNPs can also be tested by relating patterns to those observed from the GWAS and in wild populations. I selected 397 of the candidate SNPs and diagnosed the mode of selection acting at the SA sites in an ABC analysis. Of these SNPs, 251 were diagnosed as being under balancing selection – more than 60% of the SNPs tested. This incidence of balancing selection is higher than was observed among the short intron SNPs in Chapter 3 (63.2% versus 44%) and the total proportion of SNPs classed as each selection model was different between the SA candidate SNPs and the short intron SNPs (χ^2_3=100.506, p<0.0001). The difference between these SA SNPs and the short intron SNPs is due to a higher number of SA loci than short intron SNPs being diagnosed as under balancing selection, as well as those where no model could be assigned (and

accordingly a lower number of directionally selected loci in SA SNPs compared to the short intron SNPs).

While this result is in line with the expectation that SA should generate additional balancing selection at candidate loci, a direct comparison between these two sets of SNPs is confounded by any differences in their genomic context. Short intron SNPs occur sandwiched between coding sequences, and the introns are typically 66bp or shorter in length (Parsch et al. 2010). In contrast, the SA SNPs are found in a range of locations across the genome including protein coding exons, regulatory regions, and inter-genic regions (Ruzicka et al. 2019). This means that SA SNPs are on average located further away than short intron SNPs from non-synonymous sites, which experience the strongest effects of purifying selection (Andolfatto 2005), and are expected to be less affected by selection at linked loci. This difference has important implications when interpreting the elevated incidence of balancing selection in SA SNPs. In Chapter 3, I suggested that the patterns of selection at short intron SNPs could be a result of associative overdominance (AO) (Ohta 1971). Given the average distance between focal SNPs and non-synonymous sites, AO is more likely to affect allele frequency change at short intron SNPs than at SA SNPs. If the diagnosis of balancing selection is confounded by AO, as is a concern in Chapter 3, it would be predicted that short intron SNPs would have higher rates of balancing selection than SA SNPs. But this is not what we observe. The high incidence of balancing selection at SA SNPs therefore plausibly reflects true balancing selection more than at short intron SNPs, which are likely affected by AO. The larger proportion of SA SNPs diagnosed as being under balancing selection is therefore more striking than the numbers alone would suggest. This result helps to corroborate the identification of SA candidates by Ruzicka et al., and adds specific evidence for the SA status of these 251 SNPs.

4.6.2 Comparison of results to those of Ruzicka et al. (2019)

I observed a positive relationship between predicted equilibrium MAF in the cage populations and the MAF in both the DGRP and DPGP3 lines (Figure 4.7B-C). What this demonstrates is that the short-term frequency changes observed in my experiments are reflective of the frequencies that alleles attain in wild populations. The patterns of MAF being consistent between the cage populations and the DGRP and DPGP3 populations indicates that the SNPs are potentially maintained by balancing selection across all three populations. This also shows that, despite differences in environment conditions or demography, the action of balancing selection is relatively consistent in maintaining these polymorphisms.

Despite the high proportion of SNPs displaying signals of balancing selection, I did not find any evidence for a link between the SA effect size calculated by Ruzicka et al. (2019) and the probability of balancing selection. A positive relationship between the two could have been expected, but only if the fitness effects on male and female fitness balanced each other out or if trait specific dominance occurs (Connallon and Chenoweth 2019), both of which were not explicitly determined by Ruzicka et al. (2019). Another, plausible explanation is that my analysis only included candidate SA SNPs. One of the conditions of a SNP being designated as a candidate was a SA absolute effect size of >0.2. All non-candidate SNPs, the vast majority of which have an effect size <0.2, were not included in my analysis. These non-candidate SNPs are far less likely to be under some form of balancing selection. When Ruzicka et al. (2019) found a relationship between SA effect size and MAF, and between effect size and Tajima's D from the DGRP and DPGP3 populations they not only compared candidate SNPs but also non-candidate SNPs. Candidate SNPs comprise only 0.3% of all the SNPs assayed by Ruzicka et al. (2019). Including non-candidate SNPs should make any pattern between effect size and balancing selection more obvious and increase the statistical power to detect a relationship. Increasing statistical power will be especially important if the relationship between the

effect size and balancing selection (Figure 4.6A) or in the inference of effect sizes by Ruzicka et al. (2019) was noisy (both of which are likely to be true). The addition of non-candidate SNPs in my analysis may have shown a relationship between the strength of SA and the probability of balancing selection.

There are other possible explanations for a lack of an observed relationship. The data used in the GWAS study came from flies of the long-term laboratory LH_M population whereas the cage populations were established using the DGRP lines. It is possible that genetic differences between the populations could result in the lack of correlation between effect size and the probability of balancing selection. SA loci are expected to be under strong selection to resolve conflict through a number of mechanisms, including: sex-specific expression (Hollis et al. 2015), splicing (Pennell and Morrow 2013), dominance (Grieshop and Arnqvist 2018; Spencer and Priest 2016), sex linkage (Rice 1984), and gene duplication (Connallon and Clark 2011; VanKuren and Long 2018). Resolution of SA in the DGRP population could result in a lack of relationship between the SA effect size of the LH_M population and the probability of balancing selection. However, the North American origin of both the DGRP and LH_M, means that there has been little time for the necessary wholesale changes in genetic architecture to evolve (Llaurens et al. 2017; Stephen and Li 2007). It's also possible that differences in conditions between the cages and the phenotypic assays contributing to the GWAS may alter the relationship between effect size and balancing selection (Connallon 2015; Connallon and Hall 2016, Wittman et al. 2017), but this seems unlikely.

Tajima's D can be used to detect sequence differences that indicate if a site is behaving neutrally or under some form of selection (Tajima 1989), where the action of balancing selection will result in elevated values of Tajima's D. Therefore, a positive relationship between Tajima's D values from the DGRP or DPGP3 populations and the probability of balancing

selection estimated in this study would support the occurrence of balancing selection at a locus. This would also mean that signatures of balancing selection detected in the cage populations are likely to reflect balancing selection in other populations. Such loci would be prime locations for balancing selection maintaining genetic variation over long periods of time. However, I did not find any relationship between the regional value of Tajima's D in the DGRP or DPGP3 populations and the probability of balancing selection in our cages, although the relationship did come close to statistically significant for the DPGP3 population (Table 4.4 and Figure 4.6B-C). These tests likely suffer from the same problems as those between the SA effect size and balancing selection. If Tajima's D increases with the strength of balancing selection, then limiting the analysis to only those sites expected to be under balancing selection will constrain the variation seen in both variables. As before, expanding the analysis to include non-candidate SA SNPs would allow greater chance that a relationship will be observed.

Overall, when accounting for confounding measures, there was no effect of the starting frequency on the final result of each locus (Figure 4.8; Table 4.4). Initially, when including all loci, a relationship was observed ($p<0.001$). However, I found that this was due to SNPs starting at low frequencies (<0.15), as estimated from their DGRP contributions, which had a very low probability of balancing selection and a predicted MAF of zero. Removing these loci from the analysis resulted in no relationship between the estimated starting frequency and the final predicted MAF (Figure 4.8A). This was also the case when only those loci with a final replicated MAF of zero were removed from the analysis (Figure 4.8B). It is possible that during the crossing of lines to establish the cages or soon after establishment when the populations were smaller, some variation may have been lost through drift, or their frequencies were reduced from that expected, leading to earlier loss of variation. Since models are evaluated based on the frequency trajectories, loci where fixation occurs are diagnosed as being under directional selection or neutrality, and these low starting frequency alleles

were more likely to be classified as such. Rather than being a detraction from the findings, the fact that low starting frequency SNPs are less likely to be diagnosed as balancing selection means that there are probably more SNPs under balancing selection than are currently reported by the analysis since low frequency SNPs are more often fixed early on.

The fate of these low starting frequency SNPs and their effect on diagnosing the selection model reveal an underlying issues with the methodology used both here and in Chapter 3. For sites other than *fruitless,* the starting frequency was not controlled for, but is calculated from the contributing proportion of the DGRP lines used to establish each cage. This assumed that each fly contributed equally to founding the cage population. The cages were established by taking 25 flies of each sex from each DGRP line and crossing these with the flies from the other lines that were fixed for the same *fruitless* allele. These populations were maintained for three generations and then flies were haphazardly selected to form the founding populations. Although the flies were maintained in standardised optimal conditions, there is a chance that selection may have occurred, altering the starting frequencies. A bigger problem is the possible occurrence of drift during this process as the relatively small group of flies from each line could result in the unequal contribution of individuals to the next generation. Since the cages were founded with only 10% frequency of one of the two *fruitless* alleles, only 50 flies were selected from one of the two *fruitless* populations. These processes could result in a disparity between the predicted allele frequency and the real founding frequency of each cage. Across SA SNP, the correlation between the estimated starting frequency and the mean allele frequency at timepoint 1 (after ~2 generations) was r=0.424, with an average absolute difference between the two measures of 0.117. Although some frequency change will have occurred between the start of the experiment and timepoint 1, the sizable difference between the two frequency estimates supports the idea that alleles with low estimated starting frequencies based on DGRP contributions could have been fixed due to drift. They would have

been categorised as directional selection in our analysis since using an inaccurate starting frequency affects the simulation modelling by starting the simulations from a false position. While this likely doesn't affect the diagnosis of the majority of loci, especially those starting with more intermediate frequencies, this is an issue for those starting closer to fixation. The inaccurate starting frequencies may also contribute to the ~10% of loci that were undiagnosed and the large errors associated with estimated selection parameters.

4.6.3 CRISPR SNPs

A possible solution to this problem of unclear starting frequencies is to perform the analysis without using estimates from the contributing DGRP genotypes. One approach is starting the analysis from timepoint 1, and using the frequencies measured then to initialise the selection models. Although this removes two generations of evolution, it does remove the issue of selection drift affecting the starting frequencies. This was the approach taken to investigate the five loci that were chosen to create gene edited lines using CRISPR. It is difficult at this time to know how much this improves the resolution of selection as none of the chosen CRISPR SNPs is included in the 397 SNPs previously analysed. However, repeating the analysis of the 397 SNPs starting from timepoint 1 in future would provide such a comparison. The approach taken for the CRISPR SNPs could also improve the number of loci that can be investigated. The main analysis used 397 SNPs, which is only 16.7% of those identified by Ruzicka et al. (2019). This was because only biallelic autosomal SNPs, which were both expected to be polymorphic from the DGRP lines, and were found to be so in sequencing output, were investigated. If the analysis was repeated using the approach used for the CRISPR SNPs there is information covering 1360 SA candidate SNPs. This increase would provide a more robust test of the findings described here, and also give a better spread of loci across the genome, rather than just two chromosome arms. Data are also available for 182 candidate SNPs which occur on the X-chromosome, which could be

analysed if the selection models were adapted to account for sex-linkage. Some studies have shown that SA variation may be more associated with the X-chromosome than autosomes (Rice 1984, Mullon et al. 2012). Although this was not found by Ruzicka et al. (2019), how these 182 SNPs evolve may be very different from the autosomal candidates and worth investigating in the future.

Of the CRISPR SNPs, I identified two, Ugt36E1 and CG8399, which showed a strong signature of balancing selection with posterior probabilities greater than 0.75 (Table 4.5). Of the loci investigated these two are the strongest candidates to be true SA loci and future empirical investigation should start with these loci. Interestingly they both appear to be involved in intercell transport (flybase 2021c). SNP CG10477 was diagnosed as evolving neutrally in the cages, and thereby the least likely to be truly SA and is not a priority for empirical investigation. For SNPs Strn-Mlck and CG34133 no selection model could be clearly diagnosed as acting at these loci. The next step for these SNPs is to conduct experiments to test for SA fitness effects in the CRISPR lines, prioritising those with the highest probability of balancing selection.

4.6.4 Conclusion

I have investigated the mode of selection acting at 397 candidate SA loci identified from a GWAS. The incidence of balancing selection was elevated among SA loci compared to intronic SNPs, and more than 60% of SA loci show probable signatures of balancing selection, consistent with their suggested SA fitness effects. These loci likely represent true SA loci since they have both an association with SA fitness effects and evidence of current balancing selection acting upon them. Identification and description of SA has traditionally been difficult, but this process shows that we can rapidly increase the number and confidence we have in candidate SA loci. No relationship was observed between the effect size estimated by the GWAS and the probability of balancing selection in cage populations. This

discrepancy could be due to a lack of consideration of non-candidate loci and inaccurate estimation of starting allele frequencies in the populations. I did however find a positive relationship between the predicted MAF of SNPs in the cages and that observed in wild populations. This is an encouraging result as it indicates that the same loci are under a similar mode of selection in the cages as they are in the wild. Using an alternative approach to estimate these starting frequencies, I investigated five loci which have been chosen to create CRISPR gene edited lines. Two of these showed signatures of balancing selection. This methodology is a work in progress but should continue to be developed to improve the detection of balancing selection at candidate loci. Combining GWAS with experimental evolution can be a powerful approach to validate candidate SA loci. This will not only increase the number of SA loci identified but help in our understanding of the effect SA loci have on various aspects of evolution, including how they contribute to maintaining genetic variation for fitness.

4.7 Figures and Tables

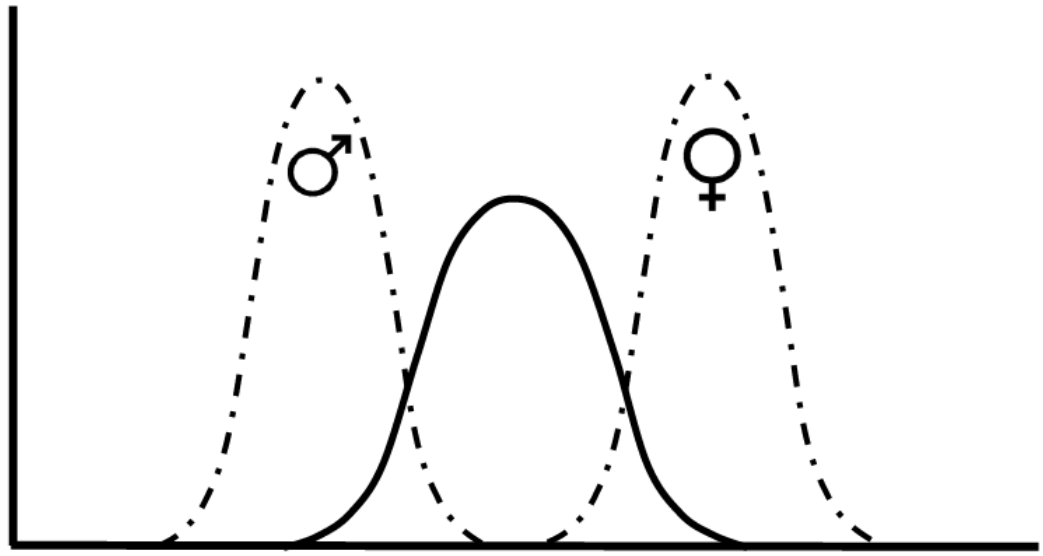


Figure 4.1. Fitness curve of trait value (x-axis) versus fitness (y-axis). The two sexes have different phenotypic fitness optima for this trait, as shown by the dashed lines. The same locus affects the trait value in both sexes resulting in sexual conflict and balancing selection. This results in the population mean for the trait (solid line) being intermediate between the sex-specific optima. The intermediate trait value results in reduced fitness for both sexes compared to the potential value if the sexes were able to adapt independently.

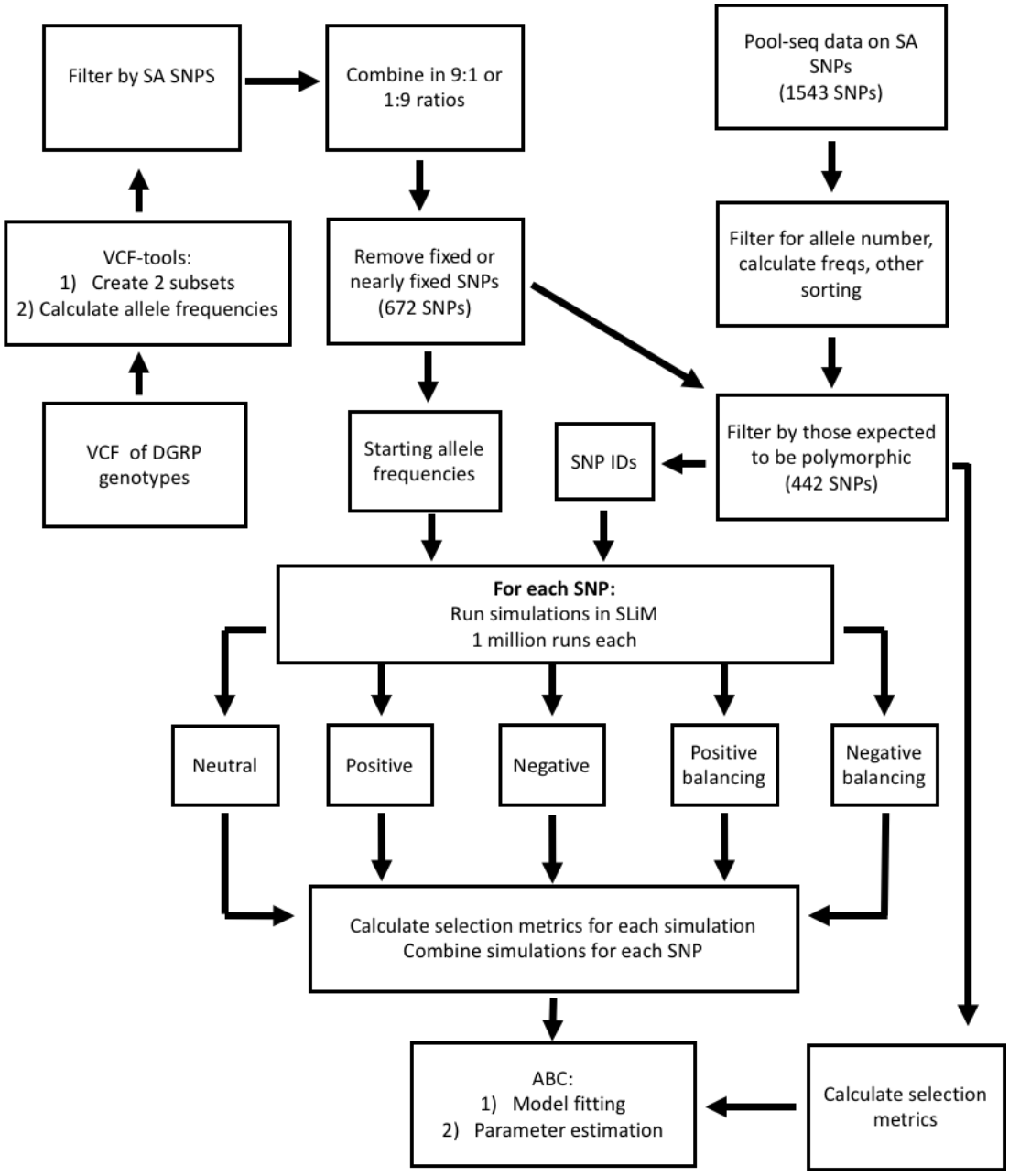


Figure 4.2. Flow diagram of the steps involved in analysing SA candidate SNPs. Starting from the VCF of the DGRPs and the _rc pool-seq data this shows step-by step the method used to calculate starting frequencies and diagnose the mode of selection acting at candidate SA SNPs reported in Ruzicka et al. (2019).

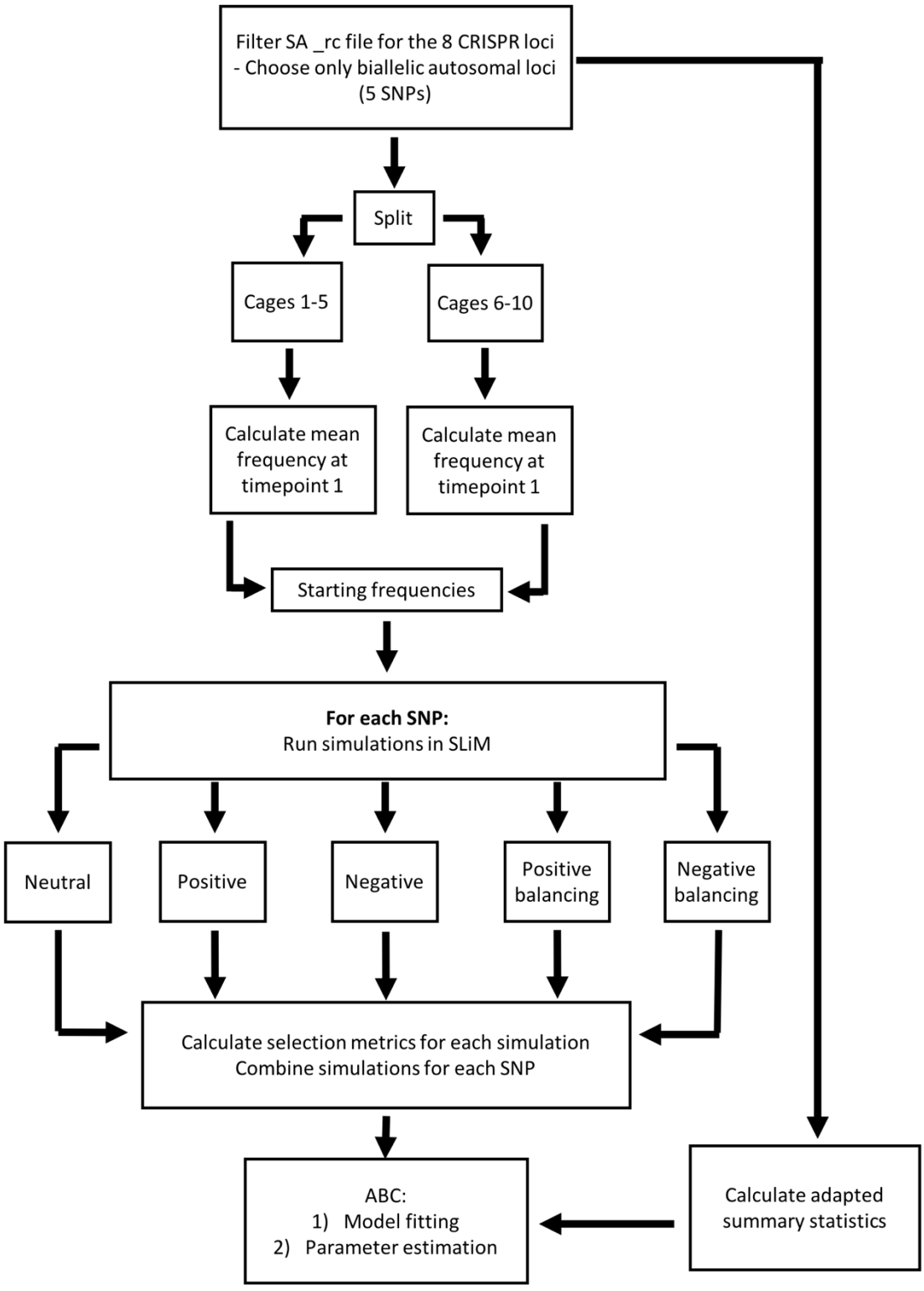


Figure 4.3. Flow diagram of steps involved in analysis of CRISPR SNPs. This is an alternative method to that of Figure 4.2. starting with the _rc file of pool-seq data to diagnose the mode of selection acting on CRISPR SNPs.

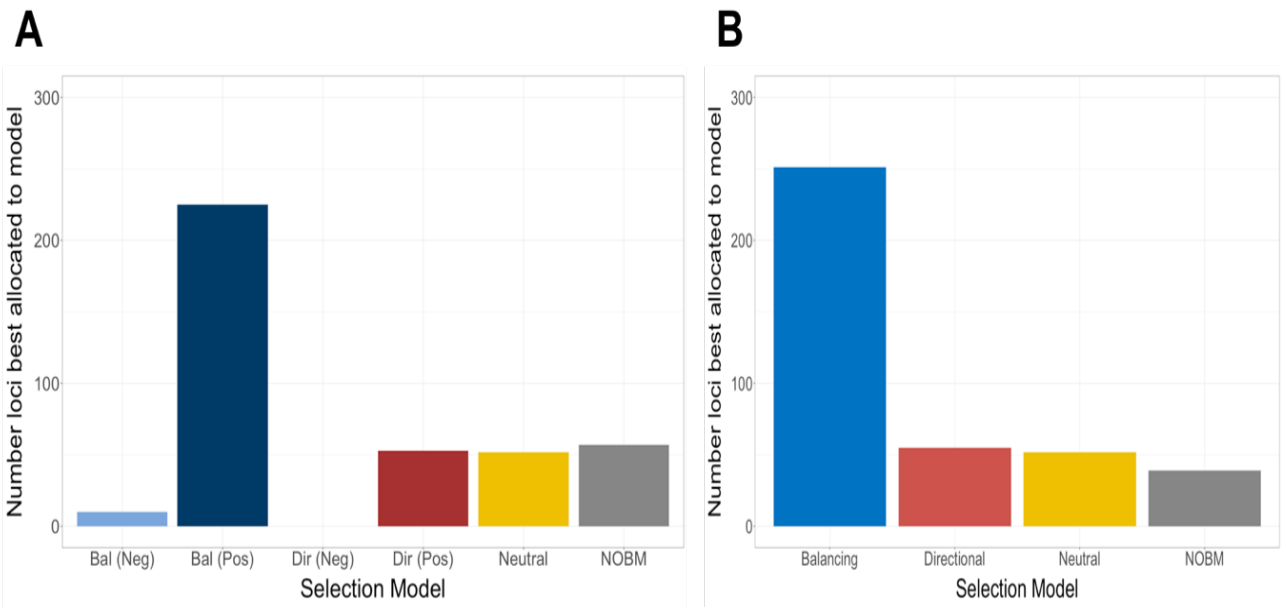


Figure 4.4. Number of SA SNPs where the posterior probability was >0.5 for each selection model. A) Counts for all five selection models, plus no overall best model (NOBM). *Positive balancing selection* (Bal (Pos)) has the highest number of loci fitting this model of selection with 56.68% of loci. No loci were diagnosed as *negative directional selection* (Dir (Neg)). B) The two forms of balancing selection and the two forms of directional selection are combined to give four probability outcomes. Balancing selection is the most commonly diagnosed selection model with 63.22% of loci.

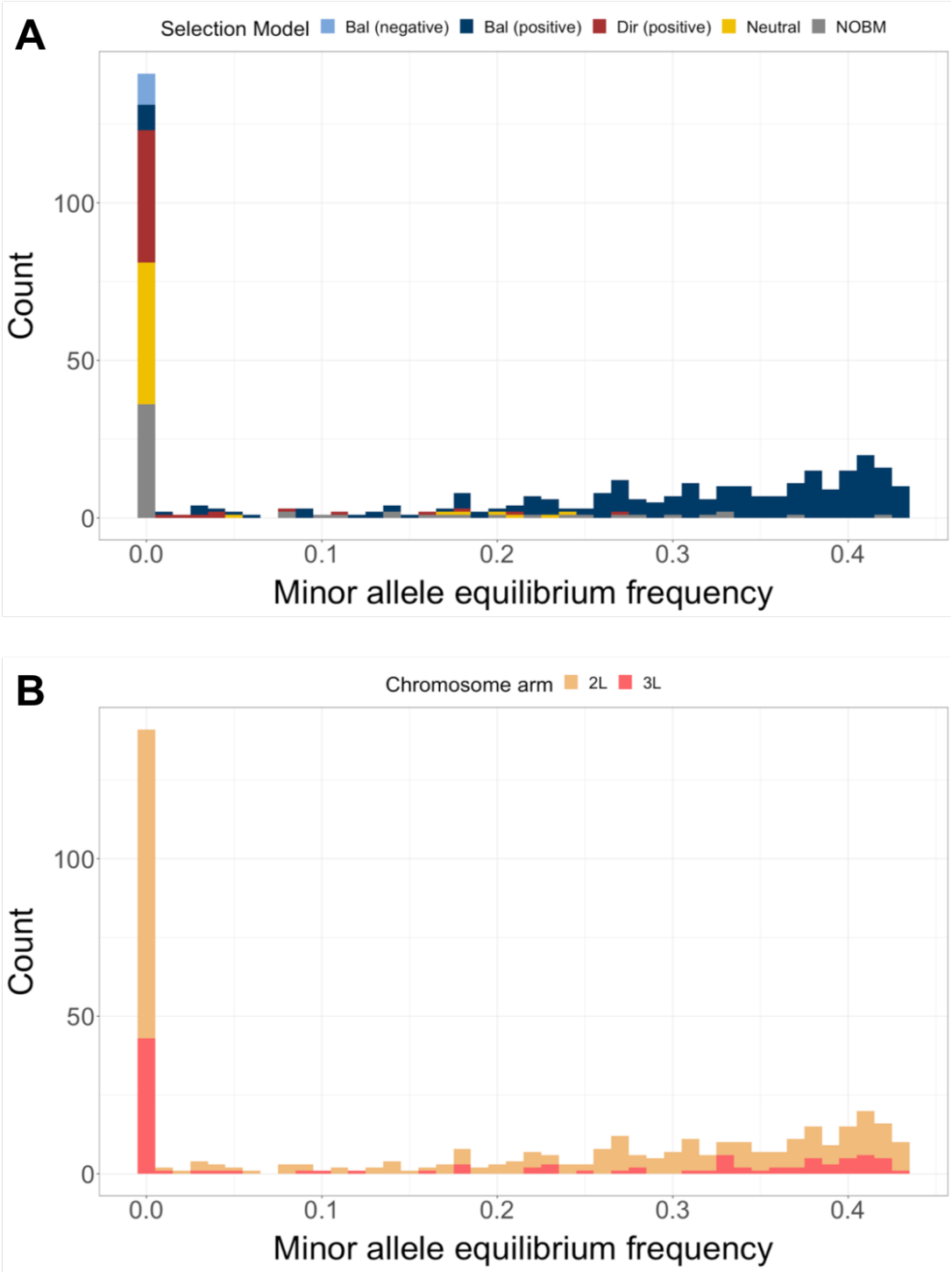


Figure 4.5. The predicted equilibrium minor allele frequency of each locus. A) Loci are coloured by the selection model diagnosed. No loci were diagnosed as *negative directional selection*. B) Loci are coloured by the chromosome arm on which they are located. There is no difference in the patterns in MAF between the chromosome arms ($F_{1,395}$=0.964, p=0.327).

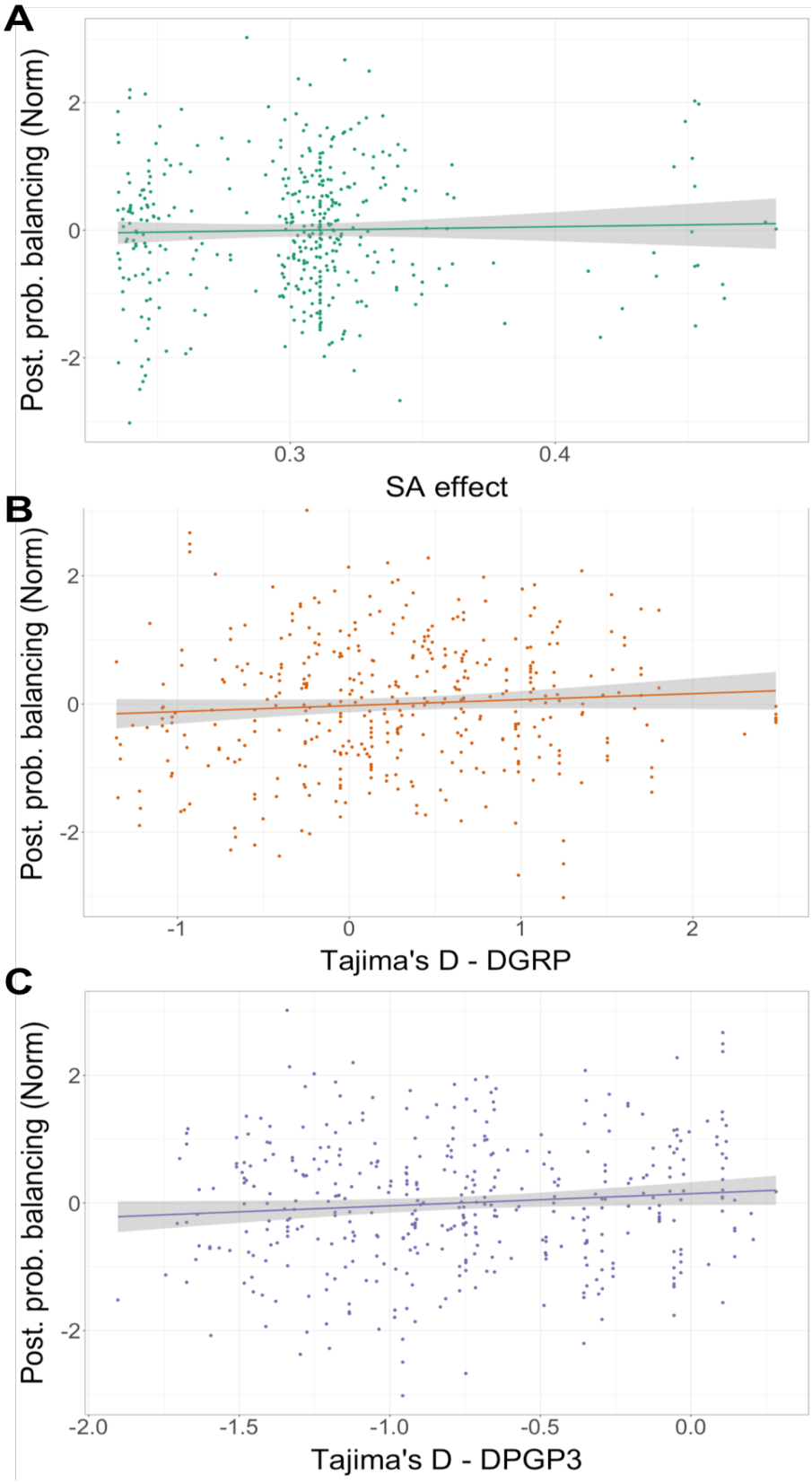

A
Post. prob. balancing (Norm)
2
0
-2
0.3
0.4
SA effect
B
Post. prob. balancing (Norm)
2
0
-2
-1
0
1
2
Tajima's D - DGRP
C
Post. prob. balancing (Norm)
2
0
-2
-2.0
-1.5
-1.0
-0.5
0.0
Tajima's D - DPGP3

Figure 4.6. Regression plots of the probability of balancing selection. In each plot the total posterior probability of balancing selection at each locus is compared against: A) SA effect size as estimated by Ruzicka et al. (2019), B) Tajima's D estimated in DGRP lines, C) Tajima's D estimated in DPGP3 lines. No pattern was observed in any of the three comparisons.

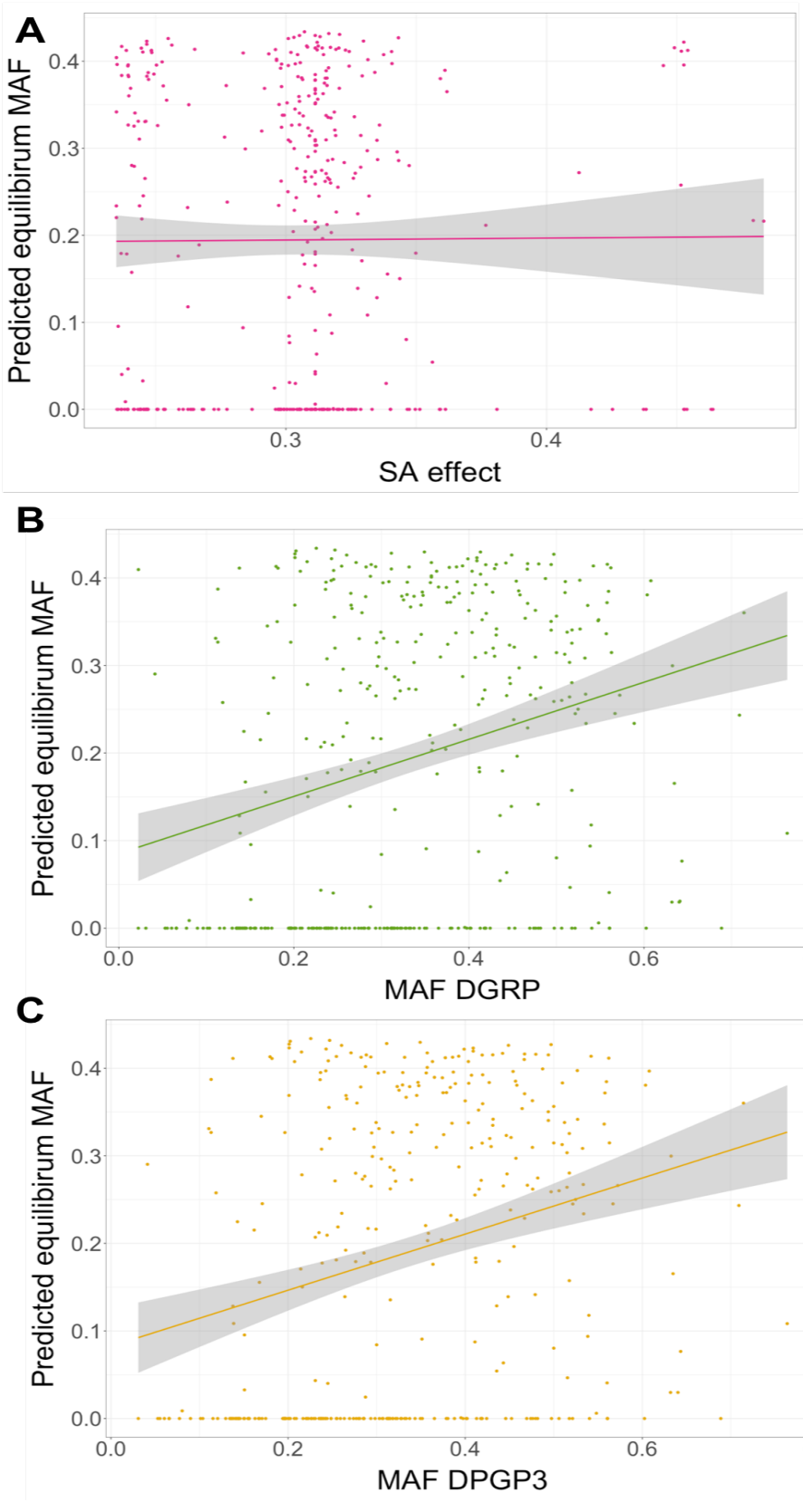

A
Predicted equilibirum MAF
0.0
0.1
0.2
0.3
0.4
0.3
0.4
SA effect
B
Predicted equilibirum MAF
0.0
0.1
0.2
0.3
0.4
0.0
0.2
0.4
0.6
MAF DGRP
C
Predicted equilibirum MAF
0.0
0.1
0.2
0.3
0.4
0.0
0.2
0.4
0.6
MAF DPGP3

Figure 4.7. Regression plots of the predicted MAF. In each plot the equilibrium MAF at each locus is compared against: A) SA effect size as estimated by Ruzicka et al. (2019), B) the equivalent MAF observed in DGRP lines, C) T the equivalent MAF observed in DPGP3 lines. No pattern was observed in A), but comparisons B) and C) show a positive relationship.

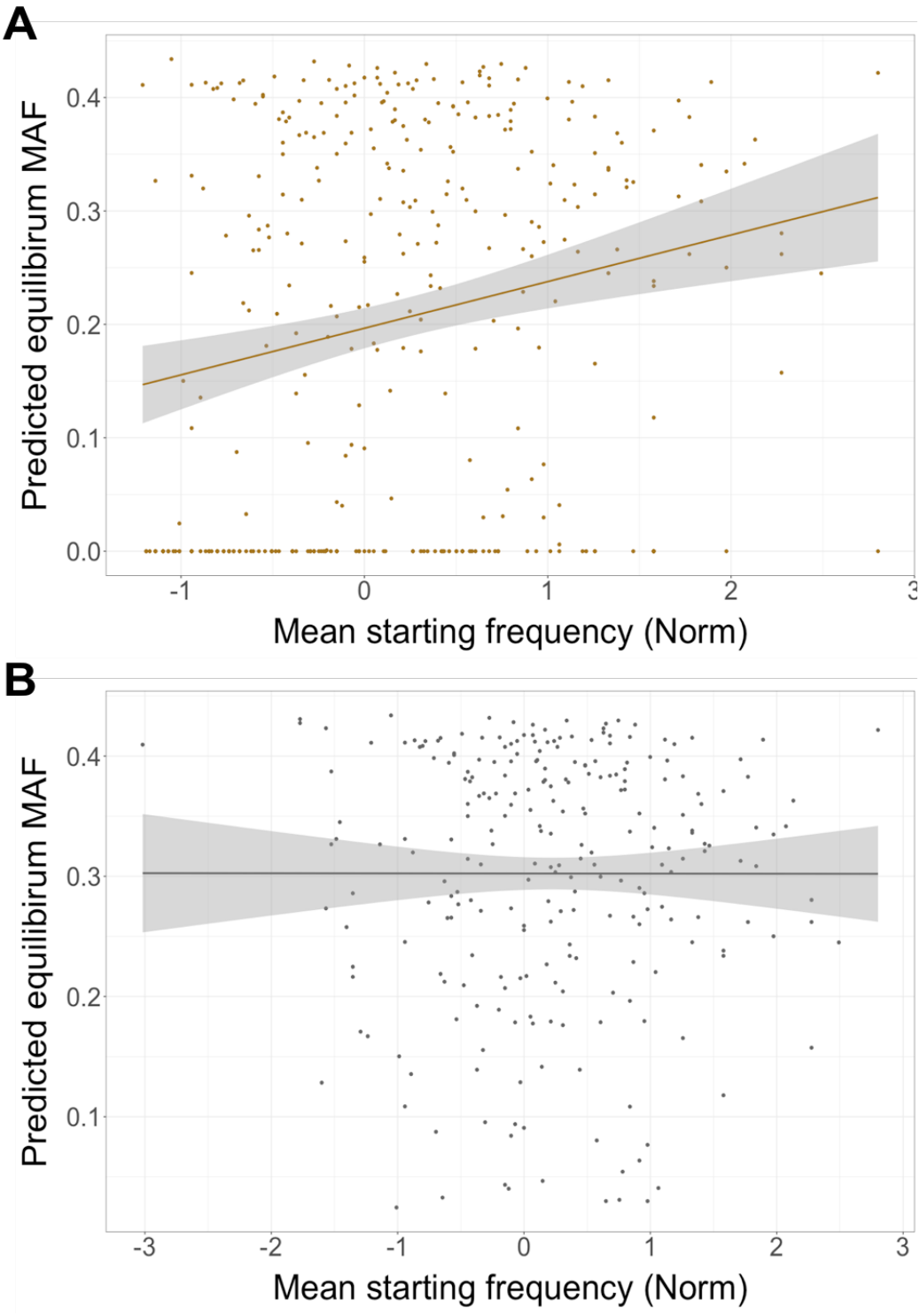


Figure 4.8. Regression plots of the predicted MAF against the estimated starting frequency. A) there is no statistical relationship despite the slight positive trend in the plot. In B) sites where MAF was predicted to be zero were removed. This resulted in no trend or relationship between MAF and starting frequency.

Chromosome	Position	Name
2L	18,827,985	Ugt36E1
2L	18,988,623	CG15170
2R	15,983,676	Strn-Mlck
2R	16,022,562	CG8399
3L	6,034,515	CG10477
3L	6,210,720	LanA
3R	29,870,043	CG34133
X	2,937,496	CG3598

Table 4.1. SNPs selected for the creation of CRISPR lines. These SNPs were identified by Ruzicka et al. (2019) as being associated with SA fitness effects. Shaded rows are biallelic autosomal SNPs and were further analysed using simulations and ABC to diagnose the mode of selection at these sites.

Gene name	Major allele	Mean freq. 1–5	Mean freq. 6–10
Ugt36E1	C	0.933	0.617
Strn-Mlck	A	0.851	0.662
CG8399	C	0.509	0.586
CG10477	A	0.561	0.658
CG34133	G	0.8	0.749

Table 4.2. Mean frequencies for each set of cages for the CRISPR SNPs. Cages 1-5 and cages 6-10 started with a different ratio of the DGRP lines (Table 3.1). A mean was taken for each set of 5 and this value used as the starting point for all further simulations. The full position numbers of the five SNPs are shown here but elsewhere these are referred to by the final 4 digits. Values are reported to 3 decimal places.

	No. of loci with >0.5 model probability		
	Total	**2L**	**3L**
Neutral	52	39	13
Positive	53	37	16
Negative	0	0	0
Balancing (positive)	225	172	53
Balancing (negative)	10	6	4
No best model	57	40	17
Neutral	52	39	13
Directional (total)	55	39	16
Balancing (total)	251	185	66
No best model	39	31	8

Table 4.3. The number of SNPs diagnosed as under each mode of selection. A SNP is classed as evolving under a particular form of selection if the posterior probability for that model is >0.5. If no probability is >0.5 then the SNP is counted in the 'no best model' category. In the top half of the table each selection model is shown separately. In the bottom half, selection types are combined to give the number of loci under that type of selection. Results are displayed for all SNPs and by the chromosome arm that they are located on.

	Explanatory variable			Chromosome arm		
Test	F	df	p	F	df	p
Balancing prob. ~ SA effect	0.231	1,394	0.631	2.024	1,394	0.155
cage MAF ~ SA effect	2.114	1,394	0.149	1.961	1,394	0.162
Balancing prob. ~ TD DGRP	3.659	1,394	0.056	0.866	1,394	0.347
Balancing prob. ~ TD DPGP3	0.024	1,394	0.877	0.886	1,394	0.347
cage MAF ~ DGRP MAF	32.609	1,390	<0.0001	0.363	1,390	0.547
cage MAF ~ DPGP3 MAF	27.78	1,359	<0.0001	0.456	1,359	0.5
cage MAF ~ starting freq.	1.225	9,337	0.278	0.545	1,337	0.461
cage MAF (>0) ~ starting freq.	0.0001	9,242	0.99	1.084	1,242	0.299
difference in starting freq. ~ cage MAF	1.274	1,390	0.26	0.963	1,390	0.327

Table 4.4. Output of linear models comparing results from cage populations to those reported by Ruzicka et al. (2019). The left-hand column indicates the test performed in each model. All models were simple linear models with one explanatory variable and chromosome arm location for each SNP. TD = Tajima's D; MAF = minor allele frequency; prob. = probability; freq. = frequency.

SNP	Model Type						
	Neutral	Positive	Negative	Balancing (positive)	Balancing (negative)	Directional (total)	Balancing (total)
Ugt36E1	0.0171	0.2153	0.0004	**0.7669**	0.0004	0.2157	**0.7673**
Strn-Mlck	0.0234	*0.4889*	0.0005	*0.4867*	0.0004	0.4894	0.4871
CG8399	0.134	0.0036	0.0954	0.1214	**0.6456**	0.099	**0.767**
CG10477	**0.8268**	0.0579	0.0311	0.0695	0.0148	0.089	0.0843
CG34133	*0.4625*	*0.4348*	0.0179	0.0791	0.0057	0.4527	0.0848

Table 4.5. Selection model probabilities for CRISPR SNPs. A SNP is diagnosed as being under a specific type of selection if the posterior probability reported for a selection model is >0.5. Values >0.5 are in bold. For SNP 3676 no model reported a probability >0.5. Instead, the two models closest are shown in italics. The probabilities of the two forms of balancing and directional selection are combined in the final two columns. Values are reported to 4 decimal places.

SNP	Selection coefficient (s)		Dominance coefficient (h)	
	Mean	95% CI	Mean	95% CI
Ugt36E1	0.078	0.007 – 0.18	1.155	0.668 – 1.638
Strn-Mlck	0.06	0.017 – 0.134	0.661	0.092 – 1.044
CG8399	0.045	0.002 – 0.118	1.652	0.067 – 3.638
CG10477	0.027	7.58e-5 – 0.166	1.333	0.002 – 4.283
CG34133	0.021	0.001 – 0.093	0.795	0.089 – 2.224

Table 4.6. Estimated parameter values of CRISPR SNPS. For each SNP the mean and 95% confidence interval are reported for the two selection parameters *s* and *h.* For SNP 3676 where $s>0$ and $h>0$ but $h<1$, this indicates that locus may be under positive directional selection. Posterior distribution size was 62500. Estimates have been converted to be positive using $s' = -s/(1+s)$ *and* $h' = 1-h$. Values are reported to 3 decimal places.

Chapter 5

5. General discussion

5.1 Overview

Genetic variation for fitness is required for selection to act and drive adaptive evolution. In this process variation decreases as variants are fixed. In spite of this, populations appear to maintain large volumes of genetic variation, and accounting for this has been a major line of inquiry for population genetics. One explanation for the maintenance of genetic variation which has received revived attention recently is balancing selection. Evidence of balancing selection comes from genomic scans which search for diagnostic signatures in variation, and from phenotypic approaches which describe the fitness impacts of polymorphisms. However, both of these approaches have their problems and are rarely used in conjunction with each other. Therefore, described polymorphisms where there is a genomic signature of balancing selection, where the fitness effects are known, and where there is evidence that the polymorphism is actively maintained by balancing selection are rare. Without such examples, our understanding of the types of loci involved and the mechanisms that maintain genetic variation by balancing selection are limited, as is our appreciation of the contribution of balancing selection to the maintenance of genetic variation for fitness more generally.

In this thesis, I have addressed several of these knowledge gaps. In this final discussion, I first summarise the main results of each chapter and then proceed to discuss these findings in the wider context of research into balancing selection and additional insights from the work developed here. I finish by describing a number of avenues for future research including developments and applications of the work presented in this thesis

5.2 Summary of main results

In the first data chapter, Chapter 2, I investigated the fitness effects of two alleles, L and S, of a *fruitless* (*fru*) indel polymorphism. This polymorphism had previously been detected at intermediate frequencies in two separate

populations of *D. melanogaster*, one in the species' ancestral range in Africa and another in the introduced range in North America. The presence of stable, worldwide polymorphism suggests that balancing selection might be occurring at this locus. Assaying the fitness of specific genotypes, I tested for any fitness differences caused by the L or S allele. I found that the S allele was positively associated with reproductive fitness traits, as females laid more eggs and males secured more mating when they inherited the S allele. However, assays of non-reproductive traits told another story with larvae inheriting the L allele more likely to survive to adulthood than S allele larvae. For lifespan there was no overall effect of either allele, but interactions between the L *fru* allele, sex and the chromosomal complement of the flies were detected. The relationship between the *fru* polymorphism and fitness, where different alleles were fittest at different life history stages is an example of antagonistic pleiotropy (AP) and could potentially generate balancing selection. I also observed evidence of epistasis and dominance reversals affecting lifespan and male mating competition traits.

To experimentally test if the *fru* polymorphism is actively maintained by balancing selection, in Chapter 3, I tracked the frequency of the *fru* alleles over 56 generations in 10 parallel populations of flies. Populations were started with different frequencies of the S and L alleles, to increase the chance of detecting any balancing selection. Sequencing allowed the construction of frequency trajectories which were compared to simulations of the experiment under 5 selection models using approximate Bayesian computation. This approach found that three SNPs just upstream of the indel’s position all reported a high probability of being under balancing selection. This suggests that the *fru* polymorphism is also likely to be under balancing selection, but the result is complicated by the fact that 44 of 100 putatively short intron SNPs were also diagnosed as under balancing selection. This could indicate that a) signatures of balancing selection are more widespread in the *D. melanogaster* genome than thought or b) there are issues with the experimental design; or c) selection patterns are affected

by associative overdominance (AO), which affects the interpretation of balancing selection at *fru*.

In Chapter 4 I investigated the mode of selection acting at 397 candidate SA loci identified by Ruzicka et al. (2019). These loci were expected to be under balancing selection due to their associated SA fitness effects and past population genetic analysis. Using time series sequencing data from the same populations as Chapter 3, I tracked SNPs' frequencies and again I employed Approximate Bayesian Computation to analyse the patterns of frequency change at these SNPs compared to simulations. More than 60% of SNPs were diagnosed as under balancing selection and are thereby the candidates where there is the greatest confidence of SA. The location of these SNPs means they are likely be less affected by the possible AO observed in Chapter 3. I found no relationship between the probability of balancing selection and the SA effect size estimated by Ruzicka et al. although this may be due to the small number of SNPs investigated. However, there was a positive relationship between the equilibrium allele frequency estimated for the cage populations and that observed in two wild populations. This provides a counter to the concerns of AO in chapter 3 as patterns of selection in the cages are consistent with patterns of selection in the wild. Finally, I developed a modified approach to diagnose selection in five loci that have been chosen to create CRISPR gene-edited lines, where no information on starting frequencies was available. Two of these show strong signals of balancing selection and should be prioritised for future fitness assays.

5.3 Balancing selection and maintaining genetic variation

In Chapter 1, I discussed the long-standing issue in population genetics of accounting for how genetic variation for fitness is maintained. The development of sequencing technology has built on earlier studies of molecular variation and showed that many species harbour large amounts of

genetic variation for traits and fitness. This has helped spur a renaissance of research into balancing selection as a potentially major mechanism of maintaining genetic variation. In this thesis I have provided evidence that balancing selection likely plays some role in maintaining polymorphisms in *D. melanogaster*. I now discuss my findings in the wider context of studies of balancing selection.

5.3.1 Methods to detect balancing selection

An issue that restricted widespread support for balancing selection in the past was the limited amount of evidence for its occurrence. This was addressed by the application of genomic scans that have identified hundreds of sites with signatures of balancing selection (Andrés et al. 2009; Croze et al. 2017; Key et al. 2014). Phenotypic studies have also contributed, now that demonstrating causative links between genetic and fitness variation are easier (Mérot et al. 2020, Glaser-Schmidt et al. 2021). Both types of studies have greatly advanced our understanding of the role that balancing selection plays in maintaining genetic variation, and they continue to do so. However, neither can demonstrate that a polymorphism is actively maintained by balancing selection (Fijarczyk and Babik 2015). I set out to rectify this problem in Chapters 3 and 4 using experimental evolution to track allele frequencies over time. This has been a moderate success, both for investigating selection at *fru* and at other loci, although it's clear that the method requires further development. In Chapter 4 specifically, the combination of GWAS and experimental evolution is a potentially powerful method to acquire experimental evidence of balancing selection occurring at candidate sites. No study has explicitly tried to detect balancing selection using this approach before, but some have observed patterns of possible balancing selection in experimental evolution studies in other species (Chelo and Teotónio 2013; Kazancıoğlu and Arnqvist 2014).

However, the application of experimental evolution to demonstrate the action of balancing selection still has its limitations. In Chapter 3, the short

intron SNPs demonstrate the confounding effect AO could have for detecting balancing selection using experimental evolution. Short intron SNPs appear to behave neutrally in *Drosophila* species over long periods of time when varying mutational patterns are accounted for (Clemente and Vogl 2012; Parsch et al. 2010), but their position near protein coding exons means that they are also more prone to the effects of linked selection such as AO. Until the amount of AO observed in Chapter 3 can be accounted for and the issue of how to differentiate it from true balancing selection is resolved, the ability to study balancing selection using experimental evolution is limited. A recent study used a machine learning approach to decipher between short-term balancing selection and incomplete selective sweeps (Isildak et al. 2021), and a similar approach could help distinguish between AO and balancing selection. The application of machine learning programs to solve problems in genetics has increased remarkably in the past few years and this could be another case where it proves useful (Schrieder and Kern 2018). I am also currently working to develop a model in SLiM (see section 5.4.2) to distinguish between AO and true balancing selection in my experiment. Alternatively, developments in sequencing technology such as longer reads and the construction of haplotype blocks could help separate AO from true balancing selection (see section 5.4.1). Although using experimental evolution to diagnose balancing selection has its issues, it's clear there are several possible solutions to help resolve them.

Different approaches have their own pros and cons, and no one approach is always going to provide the best answer to all questions (Ruzicka et al. 2020). Where phenotypic assays or experimental evolution are impractical or morally unjustifiable, such as in humans, genomic approaches are going to remain the best course of investigation. To further explore the role of balancing selection in such species, the focus should be on collecting more, and better quality, genetic data. Large genetic and phenotypic databases such as the UK Biobank (Bycroft et al. 2018) will be highly useful for this and studies have begun to utilise these large databases

to look for evidence of antagonistic polymorphisms in humans (Ruzicka et al. 2021).

5.3.2 Wider trends of balanced polymorphisms

We must be careful that the hunt for examples of balancing selection does not descend into an exercise in collection for its own sake. The focus should not just be on collecting more described phenotypic polymorphisms, or on the number of polymorphisms detected by genomic scans, but on what these can actually tell us about balancing selection. The goal is ultimately to be able to use such studies to explain how and why genetic variation is the way that it is and to understand the mechanisms and contribution of balancing selection. To this end, the results from this thesis contribute to two emerging trends that have been found regarding the properties of balanced polymorphisms.

The first trend is dominance reversal. Dominance reversal, or context dependent dominance, is where the dominance of an allele varies between different traits, and/or sexes (Grieshop et al. 2021). In Chapter 2, I suggest that the patterns observed for some fitness traits supports the presence of dominance reversal. When assaying male competitiveness, L/S and S/S males had the same chances of securing a mating, whereas when the alleles were expressed in a hemizygous state, the S/- males had greater mating success than L/- males. For this trait the S allele appears to be dominant. The opposite occurs for larval survival. L/S individuals had nearly identical chances of survival as the L/- larvae, but S/- and S/S larvae had poorer chances of survival. For each trait the allele associated with greater fitness is the dominant allele, with the deleterious allele sheltered from selection. The fitness consequences of antagonism are thereby mitigated somewhat by trait specific dominance. Dominance reversal had previously been suggested as a mechanism to resolve or stabilise antagonism, but cases illustrating its action were rare enough for it not to be considered overly important (Curtsinger et al. 1994; Hedrick 1999). However, Grieshop et al. (2021)

recently identified nine studies (including the published results of Chapter 2 – see Appendix B) that report at least some role of dominance reversals in the maintenance of polymorphisms. Combined with new theoretical models, which show that even small amounts of context-dependent variation in dominance can help stabilise polymorphisms (Connallon and Chenoweth 2019; Grieshop and Arnqvist 2018), this suggests that dominance reversal may be a common feature of balanced polymorphisms (Grieshop et al. 2021). In such cases, the locus is still antagonistic, but the negative effects on fitness and the build-up of load that results from balancing selection are reduced, due to the recessivity of the deleterious allele in each context. This can extend the lifespan of the polymorphism by partial mitigation of its negative impact. It will be interesting in the future to see if this tendency for dominance reversal is a consistent feature of balanced polymorphisms.

The second trend seen in this thesis and in a recent study by Glaser-Schmidt et al. (2021) is the identification of balanced polymorphisms within the regulatory regions of pleiotropic genes. Previous studies of balanced polymorphisms have typically focussed on protein coding regions and are sometimes restricted to them (Fijarczyk and Babik 2015). However, the polymorphisms described in the genes *fru* (Chapters 2 and 3) and f*ezzik* (Glaser-Schmidt et al. 2021) show that the locus under selection can be located in regulatory regions. These polymorphisms may affect the regulation and expression of their associated genes and thereby impact multiple traits due to pleiotropy. Balancing selection can result as even small changes in the expression of pleiotropic genes can have knock on effects for many traits, which means there is a greater potential for antagonistic fitness effects to occur. Antagonistic polymorphisms resulting from pleiotropic genes are more difficult to resolve due to the effects resolution can have on other traits. This has been proposed as a reason why sex-specific expression is less common for pleiotropic genes (Mank et al. 2008). Being located in regulatory regions of pleiotropic genes may be a common feature of balanced polymorphisms since antagonism in such polymorphisms is difficult to resolve.

5.3.3 Genetic variation between disciplines

Away from the investigation of balancing selection, an issue that I have become aware of during the research for this thesis is that there is often a disconnect between different fields of research interested in genetic variation (Velland and Geber 2005). For example, my work has approached this from a population genetic angle, which aims at understanding the factors shaping allelic polymorphism. More applied fields make decisions about health, conservation, and agricultural practices based on the amount of genetic variation and are not so much concerned with why variation is present. Part of the disconnect may be due to conflation of 'genetic variation' and 'genetic variation for fitness'. A recent study (Teixeira and Huber 2021) noted that conservation research often collects genetic data of neutral variation to guide decision making regarding population health and extinction risk. The authors point out that there is no simple relationship between neutral variation and extinction risk. For example, the critically endangered Sumatran Orangutan (*Pongo abelii*) displays a level of genetic diversity comparable to much more common species despite being at much higher risk of extinction (Locke et al. 2011). Teixeira and Huber (2021) suggest that a better interdisciplinary understanding between population genetics and conservation is needed to better interpret measures of genetic variation. The wider application of results is often not the aim, or interest, of academic researchers, but research should be made available and applicable to other research fields, and even wider society. Accomplishing this is difficult, but interdisciplinary discussion between researchers interested in genetic variation, such as population geneticists and conservation biologists, would be a good start and allow for better quality, applicable research.

5.4 Future directions

The work presented here represents substantial progress in understanding the action of balancing selection in maintaining genetic variation. These

findings open further avenues of investigation that could be explored. Additionally, developments in the work published since this work began mean that some of the approaches taken in this thesis could be improved or adapted. Some aspects of future work have been discussed in the relevant chapters, but below I expand on some of these and other areas of future work.

5.4.1 Developments in pool-sequencing and E&R experiments

Pool-sequencing (pool-seq) was the single most important technique applied in this thesis as this generated the data for Chapters 3 and 4. Next-generation sequencing technologies such as pool-seq have revolutionised the field of population genetics by massively increasing the volume of sequencing data that can be gathered (Casillas and Barbadilla 2017). Unfortunately, the approach to apply pool-seq to experimental evolution studies is riddled with conflicting advice. Despite its cost effectiveness the application of pool-seq to measure genetic variation in *D. melanogaster* has been dominated by just a few large, very well-funded groups. This is a problem for others, as with any newly expanding field, the first publications are typically produced by those able to quickly mobilise people and resources. These initial publications (e.g. Futschik and Schlötterer 2010) can have a disproportionately large influence on the field as they set the tone and standard which other researchers feel obliged to follow. For example, the sequencing coverage I used here (~40X) was at the lower limit recommended by some authors (Schlötterer et al. 2014) but also much higher than that of others (Bergland et al. 2014; Tilk et al. 2019). The work presented here (Chapters 3 and 4), along with that of others working with pool-seq in *Drosophila* (Fournier-Level et al. 2019; Griffin et al. 2017) and other species (Kurland et al. 2019), are important as a wider community of people using this technique will bring different perspectives and approaches, which will improve the quality of science produced.

As sequencing technology is continuously innovating with new designs and approaches for experiments, it is a frequent source of publications (e.g. Tilk et al. 2019; Michalak et al. 2019; Otte and Schlötterer 2021). This poses a problem for evolve and re-sequence (E&R) experimental studies since these require time for evolution to occur, meaning that once samples are ready for analysis, some new useful method has been published which would have required a different approach from the very beginning. One interesting approach that has been developed since the start of the cage population experiment is described by Tilk et al. (2019). They showed that deep coverage of individual flies used to establish populations could form a reliable reference for each population. Since linkage disequilibrium takes time to break down, blocks of nucleotides from this base population are likely to be inherited together. This means that the genotype of loci with low coverage, <5 reads per locus, can be inferred using the genotype of nearby loci based on the sequence of the well covered founder population (Tilk et al. 2019). This method provides a simple way to improve the confidence in loci with low coverage, and now makes it possible to obtain accurate estimates of allele frequencies in *Drosophila* populations using sequencing coverage of 2-5 reads per locus (Dmitri Petrov, personal communication). This can be useful in experiments with multiple populations or multiple samples as after the first round of sequencing, subsequent pools can be sequenced with low average coverages, which saves money, thus allowing more samples to be included. The approach of Tilk et al. (2019) could have applied to my study, using the fully sequenced DGRP lines as a reference, as the lack of sequencing coverage of the *fru* indel itself necessitated in the use of proxy SNPs. However, this would have been difficult due to time constraints in submitting the samples for sequencing at the time, and, as discussed in Chapter 4, I have concerns about how accurate allele frequencies estimated from the mix of DGRPs are.

Other approaches have been developed along similar lines to create haplotype blocks from pool-seq data using the correlated responses of

alleles (Michalak et al. 2019). This approach could help account for the effect of hitchhiking alleles and pinpoint targets of selection in E&R studies (Barghi and Schlötterer 2019; Nuzhdin and Turner 2013). Haplotype reconstruction of pool-seq data is likely to become a major factor of E&R experiments in the near future and will greatly improve the inferences that can be made (Otte and Schlötterer 2021). Recently, a new software tool 'haplovalidate' has been published which detects and constructs haplotype blocks for timeseries pool-seq data (Otte and Schlötterer 2021). This could be useful to help account for the patterns of wide scale balancing selection observed at short intron loci. This problem could also be investigated using another recently published method which investigated patterns of signed linkage disequilibrium and how this is influenced by selection, drift and other effects (Sandler et al. 2021). This could be applied to check the influence linked loci may have on the mode of selection acting at focal SNPs, which may possibly account for the high number of short intron SNPs diagnosed as under balancing selection.

Apart from new approaches, the current experimental evolution study could be further developed. All 10 populations were maintained and flies collected every two months until September 2021 when the populations were discontinued. This time frame covers an additional 44 generations of evolution, nearly double the number analysed here. However, due to the Covid-19 pandemic, since March 2020 all populations have been kept at 18C rather than 25C. This was done to reduce the maintenance the cages required and the number of personnel accessing the laboratory at any one site. This means that there are ~8 generations of flies available under the original conditions described in Chapter 3, and an additional 36 generations of flies which experienced the lower temperature. While this isn't ideal for the consistency of the experiment, as the change in temperature and resulting change in food timing could shift the balance of selection (Bergland et al. 2014; Wittman et al. 2017), it does allow for other aspects to be investigated. Further sequencing would allow the investigation of how the *fru*

polymorphism reacts to this change in environmental conditions, and if the form of balancing selection at *fru* is robust to such changes. If so, the behaviour of the *fru* polymorphism presented here will have greater relevance for *D. melanogaster* populations in general. Balancing selection can be environmentally dependent, i.e. balancing selection occurs under condition one, but not under condition two, similar to the DDT-R locus in *D. melanogaster* (Daborn et al. 2002; Smith et al. 2011). The change in conditions could help identify such loci which are difficult to detect with typical genome scanning approaches (Fijarczyk and Babik 2015). Sequencing these populations could also test if the observed selection patterns continue, and the longer the experiment continues, the greater the potential to distinguish between the five selection models (Burke et al. 2014; Long et al. 2015). This may be especially important for the short intron sites where so many appear to be under balancing or directional selection. Subsequent generations will break down linkage disequilibrium and decrease the effect of linked selection and associative overdominance, which I could not rule out. Future sequencing of the cages would therefore still be valuable despite the unpreventable change in conditions.

Additional sequencing is not limited to lengthening the time series of samples. Further sequencing could create a denser version of the current timeseries as flies from six timepoints were not included in this analysis. Adding data from these collections would produce more complete and consistent allele frequency trajectories and improve the resolution between selection models (Long et al. 2015; Franssen et al. 2015). Also, more flies are available for many of the timepoints already included. The 48 flies included in each pool were a representative sample, and the larger that sample, the greater the likelihood that the pool resembles the population. However, additional flies are not available for all timepoints since some earlier collections contained fewer flies and others were used for PCR genotyping (Ruzicka 2018). Additionally, for all 90 samples, not all DNA extracted was used in the sequencing reported here. This means that further

Illumina reads could be added to increase sequencing depth, or alternative sequencing technologies used. One useful development would be the application of long-read sequencing. Current Illumina sequencing techniques produce reads up to 150bp. This can cause problems with mapping in highly variable regions or regions around indels (Bennet 2020; Palmeieri and Schlötterer 2009; Fijarczyk and Babik 2015), such as the *fru* polymorphism. Longer reads, up to thousands of base pairs in length, will help reduce these problems. For example, reads thousands of base pairs long were generated by Bachman et al. (2018) to map balancing selection in *Capsella grandiflora*. Longer reads will also be very useful in discerning linkage from pool-seq experiments and creating haplotype blocks as discussed above. However, high throughput sequencing of longer reads is not currently possible, but is a key aim for the future.

Aside from additional data, the current dataset already allows many questions regarding the evolution of polymorphic loci to be investigated. I have focussed on the tracking of specific polymorphisms, but it would also be good to have a genome wide measure of genetic variation to see if changes are consistent across cages and with balancing selection or with mutation-selection-drift balance. It would also be highly beneficial to develop a method that can detect short-term patterns of balancing selection to address the issues of genomic scans (Fijarczyk and Babik 2015; Siewert and Voight 2017; Charlesworth 2006). From the data collected here, we have a good idea of what balancing selection looks like in populations in the short term and this could be used to detect similar patterns at other loci. This may be possible using machine learning techniques. For example, we could ‘teach’ a program how a balanced polymorphism behaves and then ask it to search the genome for other sites which behave in a similar manner (Isildak et al. 2021). Finally, one recent study has found that the patterns of selection may depend more on the founder population composition rather than the selection pressures (Otte et al. 2021). The 10 cage populations described here start with different allele frequencies so the data collected here could provide a

test of this paper's findings. In Chapter 4 I showed that there was no relationship between a SNPs starting frequency and its probability of balancing selection, but the study of Otte et al. (2021) could provide a way to understand differences between individual cage populations and SNPs which started at low frequencies. Due to continuing developments in genomic analysis, the data presented here could possibly be re-analysed to investigate a variety of questions in the future.

5.4.2 Expanding and improving models of selection

The five selection models developed in Chapter 3 and 4 have been successful in diagnosing the mode of selection acting at polymorphisms. However, these models are far from perfect and are greatly simplified versions of the cage populations. The high number of short intron SNPs diagnosed as being under balancing selection, and the relatively low goodness-of-fit scores for the proxy SNPs in Chapter 3 (mean D=9.675, p=0.103), illustrate the need to improve the ability of the methodology to discern between selection models. Several improvements could be made to improve the selection models. The first is to include separate sexes in the model. Separate sexes are an important factor in maintaining balancing selection, especially SA (Connallon and Clark 2014a, 2014b). Not including these detracts from our ability to accurately simulate balancing selection. The lack of differentiated sexes or sex-chromosomes precluded the inclusion of the X-chromosome CRISPR SNP 2,937,496 and the investigation of 182 SA SNPs in Chapter 4. The location of SA variation on autosomal or sex chromosomes is theorised to have a major impact on SA evolution (Rice 1984).

As discussed above in section 5.3, sex and trait specific dominance can impact the maintenance of balancing selection (Connallon and Chenoweth 2019; Grieshop et al. 2021; Zajitschek and Connallon 2018). Allowing dominance to partially vary in selection models would be another step to improve the simulations and distinguish between modes of selection.

Another variable that could be investigated is the effective population size (N_e). This was set at 1000 in the simulations, an estimate designed to reduce drift but where there is still a limited population (Kimura and Ohta 1969a, 1969b) as occurs in the cage populations. But this value is just an estimate based on the approximate number of flies that I estimated to be contained in each cage. However, an N_e of 1000 is similar to that observed in other studies of *D. melanogaster* populations in the laboratory (Mueller et al. 2013). N_e is important factor for E&R studies, particularly for estimating selection parameters (Taus et al. 2017) and for the maintenance of genetic variation in general (Leffler et al. 2012; Ellegren and Galtier 2016; Wright 1931). In Chapter 4 I describe that SA candidate loci with low starting frequencies appear less likely to be diagnosed as balancing selection, possibly due to drift. Therefore, further simulations should investigate N_e in our populations to improve the estimate used and to better account for the influence of drift.

In Chapter 3, the substantial number of short intron SNPs diagnosed as under balancing selection means there is a difficulty in confidently distinguishing between true balancing selection and AO. AO, where linked detrimental mutations can cause a locus to falsely appear as though it is under balancing selection (Ohta and Kimura 1970; Ohta 1971), is an issue for many E&R studies (Schlötterer et al. 2015, Tobler et al. 2014). If AO and balancing selection result in different allele trajectories, it may be possible to distinguish between them using ABC to determine if a locus is under balancing selection or AO. I have been working with another UCL PhD student, Carl Mackintosh, to develop a SLiM model describing potential AO in the cage populations. To accomplish this a number of additional parameters have had to be added, including haplotype and linkage structure, the effect of linked mutations, and the recombination rate. These developments are still ongoing and I'm confident that simulations of such a model will soon allow us to discern between true balancing selection and AO.

5.4.3 Investigation of fitness in CRISPR lines

In Chapter 4 I paid particular attention to eight of the 2372 SA SNPs identified by Ruzicka et al. (2019) which have been chosen to create genetically modified lines using CRISPR/cas9 gene editing. All eight loci were found to be polymorphic in the 10 populations and I focussed on five of these SNPs, removing non-biallelic and X-linked loci. It would be desirable to adapt the analysis and selection models to account for such polymorphisms in future. The number of loci for which there is empirical proof of SA fitness effects, and where we understand the genetic architecture underlying those traits, is very limited (Pennell and Morrow 2013). The addition of eight new fully investigated SA polymorphisms would represent a large increase in the number of loci described and would be a major development for research into SA, balancing selection and the maintenance of genetic variation.

The eight SNPs have been used to create 16 gene lines, two alternative alleles for each of the eight SNPs, using CRISPR genome editing technology. These lines were designed by Ted Morrow (Karlstad University) and Jon Harper (University of Sussex), together with my supervisor Max Reuter and Filip Ruzicka (now Monash University), during his time as a PhD student at UCL, based on the results of Ruzicka et al. (2019). Characterising these lines was to be a collaborative project, where the initial fitness assays were to be designed and performed by myself, with the subsequent data forming part of this thesis. However, various issues, notably delays caused by the Covid-19 pandemic, meant that the first CRISPR lines were only received in late July 2021. This was too late for a rigorous investigation of the SA fitness effects of these lines. It is intended that experimental fitness data will be collected on these lines in the near future. The methodology described in Chapter 2 provides a good starting point for designing fitness assays and could be replicated for the CRISPR lines. Sex-specific fitness assays, such as fecundity and male competition, would be essential to investigating the sex-specific effects associated with these loci from Ruzicka et al. (2019). Assays that cover different life-history stages and are replicated for all three

genotype complements (i.e. both homozygotes and the heterozygote (Hawley and Gilliland 2006)) would allow investigation of other genetic conflicts that give rise to balancing selection, AP and overdominance (Hedrick 2012, Williams 1957). Conducting fitness assays can be laborious and time consuming (Pardo-Diaz et al. 2015; Wilkinson et al. 2015) – one reason why few balanced polymorphisms have been thoroughly investigated. However, the use of time saving approaches such as the QuantiFly program (Waithe et al. 2015) will improve this. The results of Chapter 4 indicate that future research should focus on the SNPs located at positions 2L:18,827,985 and 2R:16,022,562. They had the highest probability of balancing selection and therefore are the most likely to display strong SA fitness patterns.

An additional application of CRISPR/Cas technology could prove useful in extending the analysis from Chapter 2. In this chapter I crossed the hemiclonal *fru* lines, fixed for either S or L, with the balancer/deletion stock *Df(3R)fru*$^{4\text{-}40}$*/TM6B* stock. Genotyping of this stock found that the TM6B chromosome carries the S allele. This meant that offspring from crosses between hemiclonal lines and *Df(3R)fru*$^{4\text{-}40}$*/TM6B* stock were either hemizygous S/- or L/-, heterozygous S/L, or homozygous S/S at the *fru* locus. Since the L/L genotype was not created, no fitness measure of this genotype could be taken. While this doesn't affect the conclusion of AP, the lack of this final genotypic complement does hinder the interpretation of other results relating to epistatic effects and trait specific dominance (Hawley and Gilliland 2006). This would be interesting to investigate given the role that variable dominance has been proposed to play in mechanisms maintaining genetic variation (see section 5.3 above) (Greishop et al. 2021; Connallon and Chenoweth 2019). Without L/L genotype flies this hypothesis is difficult to fully investigate. This could be solved using CRISPR. Crossing the *fru* lines with the *Df(3R)fru*$^{4\text{-}40}$*/TM6B* stock occurred because pilot experiments showed that the fitness of the homozygous *fru* lines was affected by inbreeding depression during the creation of the lines. Lines created by

CRISPR would not suffer from this problem as the S and L alleles could be inserted into identical backgrounds.

5.4.4 Outside the laboratory: balancing selection in nature

Drosophila species have been used for genetic research for over 100 years and are now the model species of choice for thousands of research groups (Morgan 1910; Hales et al. 2015). However, most of what is known about this species comes from studying *Drosophila* in the laboratory, with comparatively limited knowledge of the ecology of the flies in the wild (Keller 2007; Mansourain et al. 2018; Markow 2015). The consequence is that findings from the laboratory may not directly translate to the wild (Mallet 2006, Markow 2015). While investigation of wild *Drosophila* populations themselves has been limited, information gathered from the wild has influenced research performed in the laboratory (Hales et al. 2015; Flatt 2021). Sampling wild populations of *Drosophila* goes back to the first molecular measures of genetic variation (Lewontin and Hubby 1966; David and Capy 1988). Regular sampling of wild populations has played a major role in genetic research and our understanding of *D. melanogaster* evolution, including the establishment of standardised genetic populations and resources such as Canton-S and Oregon R (Lindsley and Grell 1968). It has furthered understanding in varied contexts including, demographic history (Stephan and Li 2007; Arguello e al. 2019), speciation (Mallet 2006; Lachaise and Silvain 2004), adaptation to environmental change (Adrion et al. 2016) and the evolution of *Drosophila* as a human commensal (Mansourian et al. 2018; Keller 2007). Studying balancing selection in the wild provides a substantial challenge, but developments in sequencing technology and detection methods (Cheng and DeGiorgio 2020) mean that it should be a fruitful one, and crucial to our understanding of how genetic variation is maintained.

In September 2019 I took part in collecting *D. melanogaster* from an apple orchard in Kent. This was part of the biannual sampling of wild *Drosophila* on behalf of DrosEU (https://droseu.net), a consortium of

population geneticists from across Europe working with *Drosophila*. DrosEU has carried out sampling twice a year for several years in order to build a picture of the genetic makeup of Europe's *Drosophila* populations. By many laboratories collaborating, a wider range of populations can be sampled than would be possible for a single group working alone. Initiatives like DrosEU bring together the expertise of different researchers to build an invaluable resource which can be applied to investigate many population genetics questions. One question that we can ask is: do we see any sign of balancing selection in wild *D. melanogaster* populations? Initial results indicate yes (Kapun et al. 2021), with signs of seasonal allele fluctuations, similar to those observed in North American populations (Bergland et al. 2014; Machondo et al. 2021). The detection of balancing selection by AP or SA will be more difficult since this will need to be visible on top of seasonal allele fluctuations. However, by expanding the approach that I have used here, and by accounting for seasonal variation, migration, and unknown population structure this should be feasible. The detection of intrinsic balancing selection in wild populations would represent a major development in our understanding of the contribution that balancing selection makes to maintaining genetic diversity in wild populations. If the goal of population genetics is to be able to explain why and how life is the way it is, then taking what we have learnt in the laboratory and from theory and applying this to the real world should be our ultimate goal.

References

1000 Genomes Project Consortium. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*. 491, 56–65.

Abbott, JK. & Morrow, EH. (2011). Obtaining snapshots of genetic variation using hemiclonal analysis. *Trends in Ecology & Evolution*. 26(7), 359–368.

Abdul-Rahman, F, Tranchina, D. & Gresham, D. (2021). Fluctuating environments maintain genetic diversity through neutral fitness effects and balancing selection. *Molecular Biology and Evolution*. 38(10), 4362–4375

Adrion, J.R, Hahn, MW. & Cooper, BS. (2015). Revisiting classic clines in *Drosophila melanogaster* in the age of genomics. *Trends in Genetics*. 31(8), 434–444.

Agrawal, AF. & Whitlock, MC. (2012). Mutation load: the fitness of individuals in populations where deleterious alleles are abundant. *Annual Review of Ecology, Evolution, and Systematics*. 43, 115-135.

Allison, AC. (1954). Protection afforded by sickle-cell trait against subtertian malarial infection. *British Medical Journal.* 4857, 290–294.

Allen, JA. & Clarke, BC. (1984). Frequency dependent selection: homage to EB Poulton. *Biological Journal of the Linnean Society*. *23*(1), 15–18.

Amambua-Ngwa, A, Tetteh, KK, Manske, M, Gomez-Escobar, N, Stewart, LB, Deerhake, ME, … & Janha, O. (2012). Population genomic scan for candidate signatures of balancing selection to guide antigen characterization in malaria parasites. *PLoS Genetics*. 8, e1002992.

Anand, A, Villella, A, Ryner, LC, Carlo, T, Goodwin, SF, Song, HJ, … & Taylor, B.J. (2001). Molecular genetic dissection of the sex-specific and vital functions of the *Drosophila melanogaster* sex determination gene fruitless. *Genetics.* 158(4), 1569–1595.

Andolfatto, P. (2005). Adaptive evolution of non-coding DNA in *Drosophila*. *Nature*. 437, 1149–1152.

Andrés, AM, Hubisz, MJ, Indap, A, Torgerson, DG, Degenhardt, JD, Boyko, AR, … & Clark, AG. (2009). Targets of balancing selection in the human genome. *Molecular Biology and Evolution.* 26(12), 2755–2764.

Andrés, AM, Dennis, MY, Kretzschmar, WW, Cannons, JL, Lee-Lin, SQ, Hurle, B, … & Nielsen, R. (2010). Balancing selection maintains a form of ERAP2 that undergoes nonsense-mediated decay and affects antigen presentation. *PLoS Genetics.* 6(10), e1001157.

Arguello, JR, Laurent, S. & Clark, AG. (2019). Demographic history of the human commensal *Drosophila melanogaster*. *Genome Biology and Evolution*. 11(3), 844–854.

Arnqvist, G, Vellnow, N. & Rowe, L. (2014). The effect of epistasis on sexually antagonistic genetic variation. *Proceedings of the Royal Society B.* 281, 20140489.

Asthana, S, Schmidt, S. & Sunyaev, S. (2005). A limited role for balancing selection. *Trends in Genetics*. 21(1), 30–32.

Bachmann, JA, Tedder, A, Laenen, B, Steige, KA. & Slotte, T. (2018). Targeted long-read sequencing of a locus under long-term balancing selection in *Capsella*. *G3: Genes, Genomes, Genetics*. 8(4), 1327–1333.

Barghi, N. & Schlötterer, C. (2019). Shifting the paradigm in Evolve and resequence studies: from analysis of single nucleotide polymorphisms to selected haplotype blocks. *Molecular Ecology.* 28, 521–524.

Barrett, RD. & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology & Evolution*. 23(1), 38–44.

Barson, NJ, Aykanat, T, Hindar, K, Baranski, M, Bolstad, GH, Fiske, P, … & Kent, M. (2015). Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature*. 528, 405–408.

Barton, NH. & Keightley, PD. (2002). Understanding quantitative genetic variation. *Nature Reviews Genetics*. 3(1), 11–21.

Bateman, AJ. (1948). Intra-sexual selection in *Drosophila*. *Heredity*. 2, 349–368.

Bates, D, Maechler, M, Bolker, B. & Walker S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software.* 67, 1–48.

Beatty, J. (1987). Weighing the risks: Stalemate in the classical/balance controversy. *Journal of the History of Biology.* 20(3), 289–319.

Beaumont, MA, Zhang, W. & Balding, DJ. (2002). Approximate Bayesian computation in population genetics. *Genetics.* 162, 2025–2035

Beaumont, MA. (2019). Approximate Bayesian computation. *Annual Review of Statistics and its Application.* 6, 379–403.

Becher, H, Jackson, BC. & Charlesworth, B. (2020). Patterns of genetic variability in genomic regions with low rates of recombination. *Current Biology*. *30*(1), 94–100.

Bergland, AO, Behrman, EL, O'Brien, KR, Schmidt, PS. & Petrov, DA. (2014). Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in *Drosophila*. *PLoS Genetics*. 10, e1004775.

Bernatchez, L. (2016). On the maintenance of genetic variation and adaptation to environmental change: considerations from population genomics in fishes. *Journal of Fish Biology*. 89(6), 2519–2556.

Bennett, EP, Petersen, BL, Johansen, IE, Niu, Y, Yang, Z, Chamberlain, CA. … & Frödin, M. (2020). INDEL detection, the 'Achilles heel' of precise genome editing: a survey of methods for accurate profiling of gene editing induced indels. *Nucleic Acids Research*. 48(21), 11958-11981.

Blum, MG. & François, O. (2010). Non-linear regression models for Approximate Bayesian Computation. *Statistics and Computing.* 20, 63–73.

Bonduriansky, R. & Chenoweth, SF. (2009). Intralocus sexual conflict. *Trends in Ecology & Evolution.* 24, 280–288.

Bitarello, BD, de Filippo, C, Teixeira, JC, Schmidt, JM, Kleinert, P, Meyer, D. & Andrés, AM. (2018). Signatures of long-term balancing selection in human genomes. *Genome Biology and Evolution*. 10(3), 939–955.

Brown, KE. & Kelly, JK. (2018). Antagonistic pleiotropy can maintain fitness variation in annual plants. *Journal of Evolutionary Biology.* 31, 46–56.

Burke, MK, Dunham, JP, Shahrestani, P, Thornton, KR, Rose, MR. & Long, AD. (2010). Genome-wide analysis of a long-term evolution experiment with *Drosophila*. *Nature*. 467, 587–590.

Burke, MK, Liti, G. & Long, AD. (2014). Standing genetic variation drives repeatable experimental evolution in outcrossing populations of *Saccharomyces cerevisiae*. *Molecular Biology and Evolution*. 31, 3228–3239.

Butlin R. & Day T. (1984). The effect of larval competition on development time and adult size in the seaweed fly, *Coelopa frigida*. *Oecologia*. 63, 122–127.

Butlin R. & Day T. (1985). Genie and karyotypic selection on an inversion polymorphism in the seaweed fly, *Coelopa frigida*. *Heredity*. 54, 267–274.

Bycroft, C, Freeman, C, Petkova, D, Band, G, Elliott, LT, Sharp, K, … & Cortes, A. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 562, 203–209.

Cagliani, R, Riva, S, Biasin, M, Fumagalli, M, Pozzoli, U, Lo Caputo, S, … & Sironi, M. (2010). Genetic diversity at endoplasmic reticulum aminopeptidases is maintained by balancing selection and is associated with natural resistance to HIV-1 infection. *Human Molecular Genetics*. 19(23), 4705–4714.

Casarini, L. & Brigante, G. (2014). The polycystic ovary syndrome evolutionary paradox: A genome-wide association studies-based, in silico, evolutionary explanation. *Journal of Clinical Endocrinology and Metabolism*. 99, E2412–E2420.

Caspari, E. (1950). On the selective value of the alleles Rt and rt in *Ephestia kuhniella*. *The American Naturalist.* 84, 367–380.

Casillas, S. & Barbadilla, A. (2017). Molecular population genetics. *Genetics.* 205(3), 1003–1035.

Chapman, JR, Hill, T. & Unckless, RL. (2019). Balancing selection drives the maintenance of genetic variation in *Drosophila* antimicrobial peptides. *Genome Biology and Evolution*. 11(9), 2691–2701.

Chapman, T. (2006). Evolutionary conflicts of interest between males and females. *Current Biology*. 16, R744–R754.

Chapman, T, Arnqvist, G, Bangham, J. & Rowe, L. (2003). Sexual conflict. *Trends in Ecology & Evolution.* 18, 41–47.

Chapman, T, Liddle, L, Kalb, JM, Wolfner, MF. & Partridge, L. (1995). Cost of mating in *Drosophila melanogaster* females is mediated by male accessory gland products. *Nature*. 373, 241–244.

Charlesworth, B. (2010). Molecular population genomics: a short history. *Genetics Research*. 92, 397–411.

Charlesworth, B. (2015). Causes of natural variation in fitness: evidence from studies of *Drosophila* populations. *PNAS*. 112, 1662–1669.

Charlesworth, B. & Charlesworth, D. (2010). *Elements of Evolutionary Genetics*. Greenwood, CO: Roberts and Company Publishers.

Charlesworth, B. & Hughes KA. (2000). The maintenance of genetic variation in life-history traits, in Singh, RS. & Krimbas, CB. (eds) *Evolutionary Genetics from Molecules to Morphology*. Cambridge: Cambridge University Press, 369–392.

Charlesworth, B, Morgan, MT. & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics*, 134, 1289–1303.

Charlesworth, B, Nordborg, M. & Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetics Research*, 70(2), 155–174.

Charlesworth, B, Charlesworth, D. & Barton, NH. (2003). The effects of genetic and geographic structure on neutral variation. *Annual Review of Ecology, Evolution, and Systematics*. 34, 99–125.

Charlesworth, D. (2006). Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics*. 2(4), e64.

Charlesworth, D. & Charlesworth, B. (1980). Sex differences in fitness and selection for centric fusions between sex-chromosomes and autosomes. *Genetical Research.* 35, 205–214.

Charlesworth, D, Charlesworth, B. & Morgan, MT. (1995). The pattern of neutral molecular variation under the background selection model. *Genetics*. 141, 1619–1632.

Chelo, IM. & Teotónio, H. (2013). The opportunity for balancing selection in experimental populations of *Caenorhabditis elegans*. *Evolution*. 67, 142–156.

Cheng, X. & DeGiorgio, M. (2020). Flexible mixture model approaches that accommodate footprint size variability for robust detection of balancing selection. *Molecular biology and evolution. 37*(11), 3267–3291.

Chippindale, AK, Gibson, JR. & Rice, WR. (2001). Negative genetic correlation for adult fitness between sexes reveals ontogenetic conflict in *Drosophila*. *PNAS.* 98, 1671–1675.

Clark, AG. (1997). Neutral behavior of shared polymorphism. *PNAS*. 94, 7730–7734.

Clarke, BC. (1972). Density-dependent selection. *The American Naturalist*. 106, 1–13.

Clarke, BC. (1979). The evolution of genetic diversity. *Proceedings of the Royal Society of London B*. 205, 453–474.

Clemente, F. & Vogl, C. (2012). Unconstrained evolution in short introns? – An analysis of genome-wide polymorphism and divergence data from *Drosophila*. *Journal of Evolutionary Biology*. 25(10), 1975–1990.

Cochran, WG. (1954). Some methods for strengthening the common χ 2 tests. *Biometrics*. 10(4), 417–451.

Connallon, T. (2015). The geography of sex-specific selection, local adaptation, and sexual dimorphism. *Evolution*. 69, 2333–2344.

Connallon, T. & Chenoweth, SF. (2019). Dominance reversals and the maintenance of genetic variation for fitness. *PLoS Biology.* 17, e3000118.

Connallon, T. & Clark, AG. (2011). The resolution of sexual antagonism by gene duplication. *Genetics*. 187, 919–937.

Connallon, T. & Clark, AG. (2012). A General Population Genetic Framework for Antagonistic Selection That Accounts for Demography and Recurrent Mutation. *Genetics.* 190, 1477–1489.

Connallon, T. & Clark, AG. (2014a). Evolutionary inevitability of sexual antagonism. *Proceedings of the Royal Society B.* 281, 20132123.

Connallon, T. & Clark, AG. (2014b). Balancing selection in species with separate sexes: insights from Fisher's geometric model. *Genetics.* 197, 991–1006.

Connallon, T. & Hall, MD. (2016). Genetic correlations and sex-specific adaptation in changing environments. *Evolution*. 70, 2186–2198.

Cox, RM. & Calsbeek, R. (2009). Sexually antagonistic selection, sexual dimorphism, and the resolution of intralocus sexual conflict. *The American Naturalist.* 173, 176–187.

Crow, JF. & Kimura, M. (1970). *An Introduction to Population Genetics Theory.* Caldwell: The Blackburn Press.

Croze, M, Wollstein, A, Božičević, V, Živković, D, Stephan, W. & Hutter, S. (2017). A genome-wide scan for genes under balancing selection in *Drosophila melanogaster*. *BMC Evolutionary Biology*. 17(1), 1–12.

Csilléry, K, Blum, MG, Gaggiotti, OE. & François, O. (2010). Approximate Bayesian computation (ABC) in practice. *Trends in Ecology & Evolution.* 25(7), 410–418.

Csilléry, K, Francois, O. & Blum, MGB. (2012). abc: an R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution.* 3, 475–479.

Cui, T, Peeters, L, Pagendam, D, Pickett, T, Jin, H, Crosbie, RS, … & Gilfedder, M. (2018). Emulator-enabled approximate Bayesian computation (ABC) and uncertainty analysis for computationally expensive groundwater models. *Journal of Hydrology.* 564,191–207

Curtsinger, JW, Service, PM. & Prout, T. (1994). Antagonistic pleiotropy, reversal of dominance, and genetic polymorphism. *The American Naturalist.* 144, 210–228.

Daborn, PJ, Yen, JL, Bogwitz, MR, Le Goff, G, Feil, E, Jeffers, S, … & Feyereisen, R. (2002). A single P450 allele associated with insecticide resistance in *Drosophila*. *Science*. 297, 2253–2256.

Danecek, P, Auton, A, Abecasis, G, Albers, CA, Banks, E, DePristo, MA, … & McVean, G. (2011). The variant call format and VCFtools. *Bioinformatics.* 27(15), 2156–2158.

Darwin, C. (1859). *On the Origin of Species.* London: John Murray.

David, JR. & Capy, P. (1988). Genetic variation of *Drosophila melanogaster* natural populations. *Trends in Genetics*. 4(4), 106–111.

Day, T, Dawe, C, Dobson, T. & Hillier, P. (1983). A chromosomal inversion polymorphism in Scandinavian populations of the seaweed fly, *Coelopa frigida*. *Hereditas*. 99(1), 135–145.

Delph, LF, Andicoechea, J, Steven, JC, Herlihy, CR, Scarpino, SV. & Bell, DL. (2011). Environment-dependent intralocus sexual conflict in a dioecious plant. *New Phytologist.* 192, 542–552.

Demir, E. & Dickson, BJ. (2005). *fruitless* splicing specifies male courtship behavior in *Drosophila*. *Cell*. 121(5), 785–794.

Dempster, E. (1955). Maintenance of genetic heterogeneity. *Cold Spring Harbor Symposia on Quantitative Biology*. 20, 25–32.

Desai, MM. & Fisher, DS. (2007). Beneficial mutation–selection balance and the effect of linkage on positive selection. *Genetics*. 176(3), 1759–1798.

DeWoody, JA, Harder, AM, Mathur, S. & Willoughby, JR. (2021). The long-standing significance of genetic diversity in conservation. *Molecular Ecology.* 30(17), 4147–4154.

Ding, G, Hasselmann, M, Huang, J, Roberts, J, Oldroyd, BP. & Gloag, R. (2021). Global allele polymorphism indicates a high rate of allele genesis at a locus under balancing selection. *Heredity.* 126(1), 163–177.

Dobzhansky, T. (1955). A review of some fundamental concepts and problems of population genetics. *Cold Spring Harbor Symposia on Quantitative Biology.* 20, 1–15.

Dobzhansky, T. (1970). *Genetics of the Evolutionary Process*. New York: Columbia University Press.

Doench, JG. (2018). Am I ready for CRISPR? A user's guide to genetic screens. *Nature Reviews Genetics*. 19, 67–80.

Durmaz, E, Benson, C, Kapun, M, Schmidt, P. & Flatt, T. (2018). An inversion supergene in *Drosophila* underpins latitudinal clines in survival traits. *Journal of Evolutionary Biology*. 31(9), 1354–1364.

Ehrenreich, IM, Torabi, N, Jia, Y, Kent, J, Martis, S, Shapiro, JA, … & Kruglyak, L. (2010). Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature*. 464, 1039–1042.

Ejsmond, MJ. & Radwan, J. (2015). Red Queen processes drive positive selection on major histocompatibility complex (MHC) genes. *PLoS Computational Biology*. 11(11), e1004627.

Ellegren, H. & Galtier, N. (2016). Determinants of genetic diversity. *Nature Reviews Genetics*. 17(7), 422–433.

Elyashiv, E, Sattath, S, Hu, TT, Strutsovsky, A, McVicker, G, Andolfatto, P, … & Sella, G. (2016). A genomic map of the effects of linked selection in *Drosophila*. *PLoS Genetics*, 12(8), e1006130.

Eyre-Walker, A. & Keightley, PD. (2007). The distribution of fitness effects of new mutations. *Nature Reviews Genetics*. 8(8), 610–618.

Fedorka, KM. & Mousseau, TA. (2004). Female mating bias results in conflicting sex- specific offspring fitness. *Nature*. 429, 65–67.

Felsenstein, J. (1976). The theoretical population genetics of variable selection and migration. *Annual Review of Genetics*. 10(1), 253–280.

Fijarczyk, A. & Babik, W. (2015). Detecting balancing selection in genomes: limits and prospects. *Molecular Ecology*. 24(14), 3529–3545.

Fisher, RA. (1930). *The genetical theory of natural selection*. Oxford: Clarendon Press

Fisher, RA. (1931). The evolution of dominance. *Biological Reviews.* 6, 345–368.

Flatt, T. (2020). Life-history evolution and the genetics of fitness components in *Drosophila melanogaster*. *Genetics*. 214(1), 3–48.

flybase.com (2021a). *Dmel\fru Flybase Gene Report* [online]. Available from: http://flybase.org/reports/FBgn0004652.html [Accessed 30 September 2021]

flybase.com (2021b). Sequence Feature: Dmel\GMR22C03 [online] Available from: http://flybase.org/reports/FBsf0000162707 [Accessed 30 September 2021]

Foerster, K, Coulson, T, Sheldon, BC, Pemberton, JM, Clutton-Brock, TH. & Kruuk, LE. (2007). Sexually antagonistic genetic variation for fitness in red deer. *Nature.* 447, 1107–1110.

Fournier-Level, A, Good, RT, Wilcox, SA, Rane, RV, Schiffer, M, Chen, W, … & Robin, C. (2019). The spread of resistance to imidacloprid is restricted by

thermotolerance in natural populations of *Drosophila melanogaster*. *Nature Ecology and Evolution*. 3, 647–656.

Fowler, K. & Partridge, L. (1989). A cost of mating in female fruitflies. *Nature*. 338, 760–761.

Franssen, SU, Nolte, V, Tobler, R. & Schlötterer, C. (2015). Patterns of linkage disequilibrium and long range hitchhiking in evolving *experimental Drosophila melanogaster* populations. *Molecular Biology and Evolution.* 32(2), 495–509.

Franssen, SU, Kofler, R. & Schlötterer, C. (2017). Uncovering the genetic signature of quantitative trait evolution with replicated time series data. *Heredity.* 118, 42–51.

Friedman, D, Magnani, J, Paranjpe, D. & Sinervo, B. (2017). Evolutionary games, climate and the generation of diversity. *PLoS ONE*. 12(8), e0184052.

Fry, JD. (2010). The genomic location of sexually antagonistic variation: some cautionary comments. *Evolution*. 64, 1510–1516.

Futschik, A. & Schlötterer, C. (2010). The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics.* 186(1), 207–218.

Gabriel, SE, Brigman, KN, Koller, BH, Boucher, RC. & Stutts, MJ. (1994). Cystic fibrosis heterozygote resistance to cholera toxin in the cystic fibrosis mouse model. *Science*. 266, 107–109.

Gailey, DA, Billeter, JC, Liu, JH, Bauzon, F, Allendorfer, JB. & Goodwin, SF. (2006). Functional conservation of the *fruitless* male sex-determination gene across 250Myr of insect evolution. *Molecular Biology and Evolution*. 23, 633–643.

Garland, T. & Rose, MR. (2009). *Experimental Evolution: Concepts, Methods, and Applications of Selection Experiments*. Berkeley: University of California Press.

Gautier, M, Foucaud, J, Gharbi, K, Cézard, T, Galan, M, Loiseau, A, … & Estoup, A. (2013). Estimation of population allele frequencies from next-

generation sequencing data: pool-versus individual-based genotyping. *Molecular Ecology. 22*(14), 3766–3779.

Gibson, JR, Chippindale, AK. & Rice. WR. (2002). The X chromosome is a hot spot for sexually antagonistic fitness variation. *Proceedings of the Royal Society London B.* 269, 499–505.

Gilbert, KJ, Pouyet, F, Excoffier, L. & Peischl, S. (2020). Transition from background selection to associative overdominance promotes diversity in regions of low recombination. *Current Biology*. 30(1), 101–107.

Gilks, WP, Abbott, JK. & Morrow, EH. (2014). Sex differences in disease genetics: evidence, evolution, and detection. *Trends in Genetics*. *30*(10), 453–463.

Gillespie, JH. & Turelli, M. (1989). Genotype-environment interactions and the maintenance of polygenic variation. *Genetics*. 121(1), 129–138.

Glaser-Schmitt. A. & Parsch. J. (2018). Functional characterization of adaptive variation within a cis-regulatory element influencing *Drosophila melanogaster* growth. *PLoS Biology*. 16, e2004538.

Glaser-Schmitt, A, Wittmann, MJ, Ramnarine, TJ. & Parsch, J. (2021). Sexual antagonism, temporally fluctuating selection, and variable dominance affect a regulatory polymorphism in *Drosophila melanogaster*. *Molecular Biology and Evolution*. 38(11), 4891–4907.

Gloss AD. & Whiteman NK. (2016). Balancing selection: walking a tightrope. *Current Biology*. 26, R73–R76.

Gloss, AD, Groen, SC. & Whiteman, NK. (2016). A genomic perspective on the generation and maintenance of genetic diversity in herbivorous insects. *Annual Review of Ecology, Evolution, and Systematics*. 47, 165–187.

Graves Jr, JL, Hertweck, KL, Phillips, MA, Han, MV, Cabral, LG, Barter, TT, … & Rose, MR. (2017). Genomics of parallel experimental evolution in *Drosophila*. *Molecular Biology and Evolution.* 34(4), 831–842.

Grieshop, K. & Arnqvist, G. (2018). Sex-specific dominance reversal of genetic variation for fitness. *PLoS Biology.* 16, e2006810.

Grieshop, K, Ho, EK. & Kasimatis, KR. (2021). Dominance reversals, antagonistic pleiotropy, and the maintenance of genetic variation. *arXiv preprint.* 2109.01571.

Griffin, PC, Hangartner, SB, Fournier-Level, A. & Hoffmann, AA. (2017). Genomic trajectories to desiccation resistance: convergence and divergence among replicate selected *Drosophila* lines. *Genetics.* 205(2), 871–890.

Habel, JC. & Schmitt, T. (2012). The burden of genetic diversity. *Biological Conservation*. 147(1), 270-274.

Haldane, JBS. (1937). The effect of variation of fitness. *The American Naturalist*. 71, 337–349.

Haldane, JBS. (1949). Disease and evolution. *Ricerca Science Supplement.* 19, 3–10

Haldane, JBS. (1957). The cost of natural selection. *Journal of Genetics*. 55, 511–524.

Haldane, JBS. & Jayakar, SD. (1963). Polymorphism due to selection of varying direction. *Journal of Genetics*. 58(2), 237–242.

Halekoh, U. & Hojsgaard, S. (2014). A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models - The R package pbkrtest. *Journal of Statistical Software*. 59(9), 1–30.

Hales, KG, Korey, CA, Larracuente, AM. & Roberts, DM. (2015). Genetics on the fly: a primer on the *Drosophila* model system. *Genetics*. 201, 815–842.

Haller, BC. & Messer, PW. (2019). SLiM 3: Forward genetic simulations beyond the Wright–Fisher model. *Molecular Biology and Evolution.* 36(3), 632–637.

Han, G, Wang, W. & Dong, Y. (2020). Effects of balancing selection and microhabitat temperature variations on heat tolerance of the intertidal black mussel *Septifer virgatus*. *Integrative Zoology*. 15(5), 416–427.

Harding, CV. & Geuze, HJ. (1993). Antigen processing and intracellular traffic of antigens and MHC molecules. *Current Opinion in Cell Biology*. 5(4), 596–605.

Harper, JA, Janicke, T. & Morrow, EH. (2021). Systematic review reveals multiple sexually antagonistic polymorphisms affecting human disease and complex traits. *medRxiv*. https://doi.org/10.1101/2020.12.16.20248300

Hawley, RS. & Gilliland, WD. (2006). Sometimes the result is not the answer: the truths and the lies that come from using the complementation test. *Genetics*. 174, 5–15.

Hedrick, AV. & Temeles, EJ. (1989). The evolution of sexual dimorphism in animals: hypotheses and tests. *Trends in Ecology & Evolution*. 4(5), 136–138.

Hedrick, PW. (1986). Genetic polymorphism in heterogeneous environments: a decade later. *Annual Review of Ecology and Systematics*. 17(1), 535–566.

Hedrick, PW. (1990). Genotypic-specific habitat selection: a new model and its application. *Heredity*, *65*(2). 145–149.

Hedrick, PW. (1998). Balancing selection and MHC. *Genetica*. 104(3), 207-214.

Hedrick, PW. (1999). Antagonistic pleiotropy and genetic polymorphism: a perspective. *Heredity.* 82, 126–133.

Hedrick, PW. (2006). Genetic polymorphism in heterogeneous environments: the age of genomics. *Annual Review of Ecology, Evolution, and Systematics.* 37, 67–93.

Hedrick, PW. (2012). What is the evidence for heterozygote advantage selection? *Trends in Ecology & Evolution*. 27(12), 698–704.

Hedrick, PW, Ginevan, ME. & Ewing, EP. (1976). Genetic polymorphism in heterogeneous environments. *Annual Review of Ecology and Systematics*. 7, 1–32.

Hermisson, J. & Pennings, P. (2005). Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*. 169, 2335–2352.

Hill, MS. (2017). The genetic basis of sexual antagonism in *Drosophila melanogaster*. PhD thesis, University College London.

Hoekstra, RF, Bijlsma, R. & Dolman, AJ. (1985). Polymorphism from environmental heterogeneity: models are only robust if the heterozygote is

close in fitness to the favoured homozygote in each environment. *Genetics Research*. 45(3), 299–314.

Hollis, B, Houle, D, Yan, Z, Kawecki, TJ. & Keller, L. (2014). Evolution under monogamy feminizes gene expression in *Drosophila melanogaster*. *Nature Communications*. 5, 3482.

Höger, AC, Ilyas, M, Stephan, W, Tellier, A, Van Der Hoorn, RA. & Rose, LE. (2012). Balancing selection at the tomato RCR3 guardee gene family maintains variation in strength of pathogen defense. *PLoS Genetics*. 8(7), e1002813.

Hoskins, RA, Carlson, JW, Wan, KH, Park, S, Mendez, I, Galle, SE, ... & Celniker, SE. (2015). The Release 6 reference sequence of the *Drosophila melanogaster* genome. *Genome Research*, 25(3), 445–458.

Huang, W, Massouras, A, Inoue, Y, Peiffer, J, Ràmia, M, Tarone, AM, ... & Mackay, TF. (2014). Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines. *Genome Research*, 24(7), 1193–1208.

Hubby, JL. & Lewontin, RC. (1966). A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. *Genetics.* 54(2), 577–594.

Hughes, AR, Inouye, BD, Johnson, MT, Underwood, N. & Vellend, M. (2008). Ecological consequences of genetic diversity. *Ecology Letters*. 11, 609–623.

Hsieh, P, Dang, V, Vollger, MR, Mao, Y, Huang, TH, Dishuck, PC, … & Sorensen, M. (2021). Evidence for opposing selective forces operating on human-specific duplicated TCAF genes in Neanderthals and humans. *Nature Communications*. 12, 1–14.

Isildak, U, Stella, A. & Fumagalli, M. (2021). Distinguishing between recent balancing selection and incomplete sweep using deep neural networks. *Molecular Ecology Resources*. https://doi.org/10.1111/1755-0998.13379

Jalvingh, KM, Chang, PL, Nuzhdin, SV. & Wertheim, B. (2014). Genomic changes under rapid evolution: selection for parasitoid resistance. *Proceedings of the Royal Society B.* 281, 20132303.

Johnston, SE, McEwan, JC, Pickering, NK, Kijas, JW, Beraldi, D, Pilkington, JG, … & Slate, JON. (2011). Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. *Molecular Ecology.* 20(12), 2555–2566.

Johnston, SE, Gratten, J, Berenos, C, Pilkington, JG, Clutton-Brock, TH, Pemberton, JM. & Slate, J. (2013). Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature.* 502, 93–95.

Kapun, M. & Flatt, T. (2019). The adaptive significance of chromosomal inversion polymorphisms in *Drosophila melanogaster*. *Molecular Ecology*. 28, 1263–1282

Kapun, M, Van Schalkwyk, H, McAllister, B, Flatt, T. & Schlötterer, C. (2014). Inference of chromosomal inversion dynamics from Pool-Seq data in natural and laboratory populations of *Drosophila melanogaster*. *Molecular Ecology.* 23(7), 1813–1827.

Kapun, M, Fabian, DK, Goudet, J. & Flatt, T. (2016). Genomic evidence for adaptive inversion clines in *Drosophila melanogaster*. *Molecular Biology and Evolution*. 33(5), 1317–1336.

Kapun, M, Barrón, MG, Staubach, F, Obbard, DJ, Wiberg, RAW, Vieira, J, ... & González, J. (2020). Genomic analysis of European *Drosophila melanogaster* populations reveals longitudinal structure, continent-wide selection, and previously unknown DNA viruses. *Molecular Biology and Evolution.* 37(9), 2661–2678.

Karasov, TL, Kniskern, JM, Gao, L, DeYoung, BJ, Ding, J, Dubiella, U, … & Barrett, LG. (2014). The long-term maintenance of a resistance polymorphism through diffuse interactions. *Nature.* 512, 436–440.

Kardos, M. & Luikart, G. (2021). The genetic architecture of fitness drives population viability during rapid environmental change. *The American Naturalist*, 197(5), 511–525.

Kassambara, A, Kosinski, M. & Biecek, P. (2019). survminer: drawing survival curves using 'ggplot2'. R package version 0.4.6

Kawecki, TJ, Lenski, RE, Ebert, D, Hollis, B, Olivieri, I. & Whitlock, MC. (2012). Experimental evolution. *Trends in Ecology & Evolution*. 27(10), 547–560.

Kazancıoğlu, E. & Arnqvist, G. (2014). The maintenance of mitochondrial genetic variation by negative frequency dependent selection. *Ecology Letters*. 17, 22–27.

Keller, A. (2007). *Drosophila melanogaster's* history as a human commensal. Current Biology. 17(3), R77–R81

Key, FM, Teixeira, JC, de Filippo, C. & Andrés, AM. (2014). Advantageous diversity maintained by balancing selection in humans. *Current Opinion in Genetics and Development*. 29, 45–51.

Kidwell, JF, Clegg, MT, Stewart, FM. & Prout T. (1977). Regions of stable equilibria for models of differential selection in the two sexes under random mating. *Genetics.* 85, 171–183.

Kimura, KI, Ote, M, Tazawa, T. & Yamamoto, D. (2005). Fruitless specifies sexually dimorphic neural circuitry in the *Drosophila* brain. *Nature*. 438, 229–233.

Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature*. 217, 624–626.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.

Kimura, M. & Crow, JF. (1964). The number of alleles that can be maintained in a finite population. *Genetics*. 49(4), 725–738.

Kimura, M. & Ohta, T. (1969a). The average number of generations until fixation of a mutant gene in a finite population. *Genetics*. 61, 763–771

Kimura, M. & Ohta, T. (1969b). The average number of generations until extinction of an individual mutant gene in a finite population. *Genetics*. 63, 701–709.

Kimura, M. & Ohta, T. (1971) *Theoretical aspects of population genetics*. Princeton: Princeton University Press.

Kirkpatrick, M. (2016). The Evolution of Genome Structure by Natural and Sexual Selection. *Journal of Heredity*. 108, 3–11.

Koenig, D, Hagmann, J, Li, R, Bemm, F, Slotte, T, Neuffer, B, … & Weigel, D. (2019). Long-term balancing selection drives evolution of immunity genes in *Capsella*. *eLife*. 8, e43606.

Kofler, R. & Schlötterer, C. (2014). A guide for the design of evolve and resequencing studies. *Molecular Biology and Evolution.* 31(2), 474–483.

Kofler, R, Pandey, RV. & Schlötterer, C. (2011). PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics.* 27(24), 3435–3436.

Kruuk, LE. (2004). Estimating genetic parameters in natural populations using the 'animal model'. *Philosophical Transactions of the Royal Society of London B*. 359, 873–890.

Kurland, S, Wheat, CW, Celorio-Mancera, Mdlp, Kutschera, VE, Hill, J, Andersson, A, … & Laikre, L. (2019). Exploring a Pool-seq-only approach for gaining population genomic insights in nonmodel species. *Ecology and Evolution*. 9(19), 11448–11463.

Lachaise, D. & Silvain, JF. (2004). How two Afrotropical endemics made two cosmopolitan human commensals: the *Drosophila melanogaster-D. simulans* palaeogeographic riddle. *Genetica*. 120, 17–39.

Lack, JB, Cardeno, CM, Crepeau, MW, Taylor, W, Corbett-Detig, RB, Stevens, KA, … & Pool, JE. (2015). The *Drosophila* genome nexus: a population genomic resource of 623 *Drosophila melanogaster* genomes, including 197 from a single ancestral range population. *Genetics*. 199(4), 1229–1241.

Lack, JB, Lange, JD, Tang, AD, Corbett-Detig, RB. & Pool, JE. (2016). A thousand fly genomes: an expanded *Drosophila* genome nexus. *Molecular biology and evolution*. 33, 3308–3313.

Lamb, AM, Walker, EA. & Wittkopp, PJ. (2017). Tools and strategies for scarless allele replacement in *Drosophila* using CRISPR/Cas9. *Fly*. 11(1), 53–64.

Lande, R. (1975). The maintenance of genetic variability by mutation in a polygenic character with linked loci. *Genetics Research*. *26*(3), 221–235.

Lande, R. (1976). Natural selection and random genetic drift in phenotypic evolution. *Evolution*. 30(2), 314–334.

Lande, R. (1980). Sexual dimorphism, sexual selection, and adaptation in polygenic characters. *Evolution.* 34(2), 292–305.

Lande, R. (1981). Models of speciation by sexual selection on polygenic traits. *PNAS.* 78(6), 3721–3725.

Lande, R. & Shannon, S. (1996). The role of genetic variation in adaptation and population persistence in a changing environment. *Evolution.* 434–437.

Langmead, B. & Salzberg, SL. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*. *9*(4), 357–359.

Leffler, EM, Bullaughey, K, Matute, DR, Meyer, WK, Segurel, L, Venkat, A, … & Przeworski, M. (2012). Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS Biology.* 10(9). e1001388

Leffler, EM, Gao, Z, Pfeifer, S, Ségurel, L, Auton, A, Venn, O, … & Donnelly, P. (2013). Multiple instances of ancient balancing selection shared between humans and chimpanzees. *Science*. 339, 1578–1582.

Levene, H. (1953). Genetic equilibrium when more than one ecological niche is available. *The American Naturalist*. 87, 331–333.

Lewontin, RC. (1974). *The Genetic Basis of Evolutionary Change.* New York: Columbia University Press.

Lewontin, RC. & Hubby, JL. (1966). A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobsura*. *Genetics*. 54(2), 595–609.

Li, H, Handsaker, B, Wysoker, A, Fennell, T, Ruan, J. Homer, N, … & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics.* 25(16), 2078–2079.

Lindsley, DL. & Grell, EH. (1967) Genetic Variations of *Drosophila melanogaster*. Carnegie Institute of Washington Publication No. 627.

Lindtke, D, Lucek, K, Soria-Carrasco, V, Villoutreix, R, Farkas, TE, Riesch, R, … & Nosil, P. (2017). Long-term balancing selection on chromosomal variants associated with crypsis in a stick insect. *Molecular Ecology*. 26, 6189–6205.

Llaurens, V, Whibley, A. & Joron, M. (2017). Genetic architecture and balancing selection: the life and death of differentiated variants. *Molecular Ecology.* 26, 2430–2448.

Locke, DP, Hillier, LW, Warren, WC, Worley, KC, Nazareth, LV, Muzny, DM, … & Mitreva, M. (2011). Comparative and demographic analysis of orang-utan genomes. *Nature*. 469, 529–533.

Lohoff, FW, Dahl, JP, Ferraro, TN, Arnold, SE, Gallinat, J, Sander, T. & Berrettini, WH. (2006). Variations in the vesicular monoamine transporter 1 gene (VMAT1/SLC18A1) are associated with bipolar I disorder. *Neuropsychopharmacology*. 31(12), 2739–2747.

Lohoff, FW, Weller, AE, Bloch, PJ, Buono, RJ, Doyle, GA, Ferraro, TN. & Berrettini, WH. (2008). Association between polymorphisms in the vesicular monoamine transporter 1 gene (VMAT1/SLC18A1) on chromosome 8p and schizophrenia. *Neuropsychobiology*. 57, 55–60.

Long, A, Liti, G, Luptak, A. & Tenaillon, O. (2015). Elucidating the molecular architecture of adaptation via evolve and resequence experiments. *Nature Reviews Genetics.* 16(10), 567–582.

Long, TA. & Rice, WR. (2007). Adult locomotory activity mediates intralocus sexual conflict in a laboratory-adapted population of *Drosophila melanogaster*. *Proceedings of the Royal Society of London B*. 274, 31053112.

Lynch, M. (2011). Statistical inference on the mechanisms of genome evolution. *PLoS Genetics*. *7*(6), e1001389.

Machado, HE, Bergland, AO, O'Brien, KR, Behrman, EL, Schmidt, PS. & Petrov, DA. (2016). Comparative population genomics of latitudinal variation in *Drosophila simulans* and *Drosophila melanogaster*. *Molecular Ecology.* 25(3), 723–740.

Machado, HE, Bergland, AO, Taylor, R, Tilk, S, Behrman, E, Dyer, K, … & Kim, B. (2021). Broad geographic sampling reveals the shared basis and environmental correlates of seasonal adaptation in *Drosophila*. *eLife.* 10, e67577.

Mackay, TF, Richards, S, Stone, EA, Barbadilla, A, Ayroles, JF, Zhu, D, … & Richardson, MF. (2012). The *Drosophila melanogaster* genetic reference panel. *Nature*. 482, 173–178.

MacPherson, A, Keeling, MJ. & Otto, SP. (2021). Coevolution fails to maintain genetic variation in a host–parasite model with constant finite population size. *Theoretical Population Biology*, 137, 10–21.

Magwire, MM, Fabian, DK, Schweyen, H, Cao, C, Longdon, B. Bayer, F. & Jiggins, FM. (2012). Genome-wide association studies reveal a simple genetic basis of resistance to naturally coevolving viruses in *Drosophila melanogaster*. *PLoS Genetics.* 8(11), pe1003057.

Mallet, J. (2006). What does *Drosophila* genetics tell us about speciation? *Trends in Ecology & Evolution*. 21(7), 386–393.

Mank, JE, Hultin-Rosenberg, L, Zwahlen, M. & Ellegren, H. (2008). Pleiotropic constraint hampers the resolution of sexual antagonism in vertebrate gene expression. *The American Naturalist*, *171*(1), 35–43.

Mansourian, S, Enjin, A, Jirle, EV, Ramesh, V, Rehermann, G, Becher, PG, … & Stensmyr, MC. (2018). Wild African *Drosophila melanogaster* are seasonal specialists on marula fruit. *Current Biology*. 28(24), 3960–3968.

Mantel, N. & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*. 22(4), 719–748.

Markow, T.A. (2015). The natural history of model organisms: the secret lives of *Drosophila* flies. *eLife*. 4, e06793.

Marraffini, LA. & Sontheimer, EJ. (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nature Reviews Genetics*. 11, 181–190.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*. 17, 10–12.

Martins, NE, Faria, VG, Nolte, V, Schlötterer, C, Teixeira, L, Sucena, É. & Magalhães, S. (2014). Host adaptation to viruses relies on few genes with different cross-resistance properties. *PNAS*. 111(16), 5938–5943.

Maynard Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.

Maynard Smith, J. & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetics Research*. 23, 23–35.

Maynard Smith, J. & Hoekstra, R. (1980). Polymorphism in a varied environment: how robust are the models? *Genetics Research*. 35(1), 45–57.

McCart, C, Buckling, A. & ffrench-Constant, RH. (2005). DDT resistance in flies carries no cost. *Current Biology*. 15, R587–R589.

McKenna, A, Hanna, M, Banks, E, Sivachenko, A, Cibulskis, K, Kernytsky, A, … & DePristo, MA. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*. 20(9), 1297–1303.

McKinley, TJ, Vernon, I, Andrianakis, I, McCreesh, N, Oakley, JE, Nsubuga, RN, … & White, RG. (2018). Approximate Bayesian computation and simulation-based inference for complex stochastic epidemic models. *Statistical Science.* 33(1), 4–18.

Merilä, J, Sheldon, BC. & Ellegren, H. (1997). Antagonistic natural selection revealed by molecular sex identification of nestling collared flycatchers. *Molecular Ecology*. 6(12), 1167–1175.

Mérot, C, Berdan, EL, Babin, C, Normandeau, E, Wellenreuther, M. & Bernatchez, L. (2018). Intercontinental karyotype–environment parallelism supports a role for a chromosomal inversion in local adaptation in a seaweed fly. *Proceedings of the Royal Society B.* 285, 20180519.

Mérot, C, Llaurens, V, Normandeau, E, Bernatchez, L. & Wellenreuther, M. (2020). Balancing selection via life-history trade-offs maintains an inversion polymorphism in a seaweed fly. *Nature Communications.* 11, 670.

Mérot, C, Berdan, E, Cayuela, H, Djambazian, H, Ferchaud, AL, Laporte, M, … & Bernatchez, L. (2021). Locally adaptive inversions modulate genetic variation at different geographic scales in a seaweed fly. *bioRxiv*. 202012.

Michalak, P, Kang, L, Schou, MF, Garner, HR. & Loeschcke, V. (2019). Genomic signatures of experimental adaptive radiation in *Drosophila*. *Molecular Ecology*. 28(3), 600–614.

Micheletti, S, Parra, E. & Routman, EJ. (2012). Adaptive color polymorphism and unusually high local genetic diversity in the side-blotched lizard, *Uta stansburiana*. *PLoS ONE*. 7(10), e47694.

Miller, DE, Cook, KR, Arvanitakis, AV. & Hawley, RS. (2016). Third chromosome balancer inversions disrupt protein-coding genes and influence distal recombination events in *Drosophila melanogaster*. *G3: Genes, Genomes, Genetics.* 6, 1959–1967.

Mitchell-Olds, T, Willis, JH. & Goldstein, DB. (2007). Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nature Reviews Genetics*. 8(11), 845–856.

Mokkonen, M, Kokko, H, Koskela, E, Lehtonen, J, Mappes, T, Martiskainen, H. & Mills, SC. (2011). Negative frequency-dependent selection of sexually antagonistic alleles in *Myodes glareolus*. *Science*. 334, 972–974.

Morgan, TH. (1910). Sex limited inheritance in *Drosophila*. *Science*. 32, 120–122.

Mueller, LD, Joshi, A, Santos, M. & Rose, MR. (2013). Effective population size and evolutionary dynamics in outbred laboratory populations of *Drosophila*. *Journal of genetics*. 92(3), 349–361.

Muller, HJ. (1950). Our load of mutations. *American Journal of Human Genetics*. 2, 111–176.

Muller, HJ. (1958). Evolution by Mutation. *Bulletin of the American Mathematical Society.* 64, 137–160.

Mullon, C, Pomiankowski, A. & Reuter, M. (2012). The effects of selection and genetic drift on the genomic distribution of sexually antagonistic alleles. *Evolution.* 66(12), 3743–3753.

Neville, MC, Nojima, T, Ashley, E, Parker, DJ, Walker, J, Southall, T, … & Russell, S. (2014). Male-specific fruitless isoforms target neurodevelopmental genes to specify a sexually dimorphic nervous system. *Current Biology*. 24(3), 229–241.

Nojima, T, Neville, MC. & Goodwin, SF. (2014). Fruitless isoforms and target genes specify the sexually dimorphic nervous system underlying *Drosophila* reproductive behavior. *Fly.* 8, 95–100.

Nolte, V. & Schlötterer, C. (2008). African *Drosophila melanogaster* and *D. simulans* populations have similar levels of sequence variability, suggesting comparable effective population sizes. *Genetics.* 178, 405–412.

Novak, S. & Barton, NH. (2017). When does frequency-independent selection maintain genetic variation? *Genetics*. 207(2), 653–668.

Novičić, ZK, Sayadi, A, Jelić, M. & Arnqvist, G. (2020). Negative frequency dependent selection contributes to the maintenance of a global polymorphism in mitochondrial DNA. *BMC Evolutionary Biology*. 20, 1–9.

Nuzhdin, SV. & Turner, TL. (2013). Promises and limitations of hitchhiking mapping. *Current Opinion in Genetics and Development*. 23(6), 694-699.

Ohta, T. (1971). Associative overdominance caused by linked detrimental mutations. *Genetics Research.* 18(3), 277–286.

Ohta, T. (1973). Slightly deleterious mutant substitutions in evolution. *Nature*. 246, 96–98.

Ohta, T. (1992). The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics*. 23(1), 263–286.

Ohta, T. & Kimura, M. (1970). Development of associative overdominance through linkage disequilibrium in finite populations. *Genetical Research*. 16, 165–177.

Ono, J, Gerstein, AC. & Otto, SP. (2017). Widespread genetic incompatibilities between first-step mutations during parallel adaptation of *Saccharomyces cerevisiae* to a common environment. *PLoS Biology.* 15, 1–26.

Orozco-terWengel, P, Kapun, M, Nolte, V, Kofler, R, Flatt, T. & Schlötterer, C. (2012). Adaptation of *Drosophila* to a novel laboratory environment reveals temporally heterogeneous trajectories of selected alleles. *Molecular Ecology*. 21(20), 4931–4941.

Orr, HA. (2009). Fitness and its role in evolutionary genetics. *Nature Reviews Genetics*. 10(8), 531–539.

Otte, KA. & Schlötterer, C. (2021). Detecting selected haplotype blocks in evolve and resequence experiments. *Molecular Ecology Resources*. 21(1), 93–109.

Otte, KA, Nolte, V, Mallard, F. & Schlötterer, C. (2021). The genetic architecture of temperature adaptation is shaped by population ancestry and not by selection regime. *Genome Biology.* 22, 1–25.

Palmieri, N. & Schlötterer, C. (2009). Mapping accuracy of short reads from massively parallel sequencing and the implications for quantitative expression profiling. *PLoS ONE*. 4(7), e6323.

Pardo-Diaz, C, Salazar, C. & Jiggins, CD. (2015). Towards the identification of the loci of adaptive evolution. *Methods in Ecology and Evolution*. 6, 445–464.

Parker, DJ, Gardiner, A, Neville, MC, Ritchie, MG. & Goodwin, SF. (2014). The evolution of novelty in conserved genes; evidence of positive selection in the *Drosophila* fruitless gene is localised to alternatively spliced exons. *Heredity*. 112(3), 300–306.

Parker, GA. (1979). Sexual Selection and Sexual Conflict, in Blum, MS. & Blum, NA. (ed) *Sexual Conflict and Reproductive Competition in Insects.* London: Academic Press, 123-166.

Parker, GA, Baker, RR. & Smith, VGF. (1972). The origin and evolution of gamete dimorphism and the male-female phenomenon. *Journal of Theoretical Biology*. 36(3), 529–553.

Parsch, J*,* Novozhilov, S*,* Saminadin-Peter, SS*,* Wong, KM. & Andolfatto, P. (2010)*.* On the utility of short intron sequences as a reference for the detection of positive and negative selection in *Drosophila. Molecular Biology and Evolution*. 27, 1226–1234*.*

Pauls, SU, Nowak, C, Bálint, M. & Pfenninger, M. (2013). The impact of global climate change on genetic diversity within populations and species. *Molecular ecology*. *22*(4), 925–946.

Penna, S, Ghag, SB, Ganapathi, TR. & Jain, SM. (2019). Induced genetic diversity in banana, in D, Nandwani. (ed) *Genetic Diversity in Horticultural Plants*. Cham: Springer Nature, 273–297.

Pennell, TM. & Morrow, EH. (2013). Two sexes, one genome: the evolutionary dynamics of intralocus sexual conflict. *Ecology and Evolution*. 3, 1819–1834.

Pennell, TM, de Haas, FJ, Morrow, EH. & van Doorn, GS. (2016). Contrasting effects of intralocus sexual conflict on sexually antagonistic coevolution. *PNAS.* 113(8), E978–E986.

Peterson, RA. (2021). Finding optimal normalizing transformations via bestNormalize. *The R Journal*. 13, 310–329.

Peterson, RA. & Cavanaugh, JE. (2020). Ordered quantile normalization: a semiparametric transformation built for the cross-validation era. *Journal of Applied Statistics*. 47, 2312–2327.

Phillips, MA, Long, AD, Greenspan, ZS, Greer, LF, Burke, MK, Villeponteau, B, … & Rose, M.R., (2016). Genome-wide analysis of long-term evolutionary domestication in *Drosophila melanogaster*. *Scientific reports*. 6(1), 1–12.

Picard Toolkit. (2019). *A set of command line tools (in Java) for manipulating high-throughput sequencing (HTS) data and formats such as SAM/BAM/CRAM and VCF* [online]. Available from: http://broadinstitute.github.io/picard/ [Accessed 11 October 2021]

Pischedda, A. & Chippindale, AK. (2006). Intralocus sexual conflict diminishes the benefits of sexual selection. *PLoS Biology*. 4, e356.

Price, DK. & Burley, NT. (1994). Constraints on the evolution of attractive traits: selection in male and female zebra finches. *The American Naturalist*. 144(6), 908–934.

Pritchard JK, Seielstad MT, Perez-Lezaun A. & Feldman MW. (1999). Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution.* 16, 1791–1798.

Prout, T. (2000). How well does opposing selection maintain variation? in Singh, RS. & Krimbas CB. (eds) *Evolutionary Genetics from Molecules to Morphology*. Cambridge: Cambridge University Press, 369–392.

Purcell, S, Neale, B, Todd-Brown, K, Thomas, L, Ferreira, MA, Bender, D, … & Sham, PC. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*. 81(3), 559–575.

R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ratan, A, Olson, TL, Loughran, TP. & Miller, W. (2015). Identification of indels in next-generation sequencing data. *BMC Bioinformatics*. 16(1), 1–8.

Reed, DH. & Frankham, R. (2003). Correlation between fitness and genetic diversity. *Conservation Biology*, *17*(1), 230–237.

Remolina, SC, Chang, PL, Leips, J, Nuzhdin, SV. & Hughes, KA. (2012). Genomic basis of aging and life-history evolution in *Drosophila melanogaster*. *Evolution*. 66(11), 3390–3403.

Rice, WR. (1984). Sex chromosomes and the evolution of sexual dimorphism. *Evolution*. 38, 735–742.

Rice, WR. (1992). Sexually antagonistic genes: experimental evidence. *Science.* 256, 1436–1439.

Rice, WR. (1996). Sexually antagonistic male adaptation triggered by experimental arrest of female evolution. *Nature*. 381, 232–234.

Rice, WR. (2000). Dangerous liaisons. *PNAS*. 97, 12953–12955.

Rice, WR. & Chippendale, AK (2001) Intersexual ontogenetic conflict. *Journal of Evolutionary Biology*. 14, 685–693.

Rice, WR, Linder, JE, Friberg, U, Lew, TA, Morrow, EH. & Stewart, AD. (2005). Inter-locus antagonistic coevolution as an engine of speciation: assessment with hemiclonal analysis. *PNAS.* 102, 6527–6534.

Richman, A. (2000). Evolution of balanced genetic polymorphism. *Molecular Ecology*. 9(12), 1953-1963.

Rideout, EJ, Billeter, JC. & Goodwin, SF. (2007). The sex-determination genes *fruitless* and *doublesex* specify a neural substrate required for courtship song. *Current Biology.* 17, 1473–1478.

Roberts, RB, Ser, JR. & Kocher, TD. (2009). Sexual conflict resolved by invasion of a novel sex determiner in Lake Malawi cichlid fishes. *Science*. 326, 998–1001.

Romiguier, J. Gayral, P. Ballenghien, M. Bernard, A. Cahais, V. Chenuil, A. … & Loire, E. (2014). Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature*. 515, 261–263.

Rose, MR. (1982). Antagonistic pleiotropy, dominance, and genetic variation. *Heredity.* 48, 63–78.

Rose, MR. (1985). Life history evolution with antagonistic pleiotropy and overlapping generations. *Theoretical Population Biology.* 28, 342–358.

Rose, MR. & Charlesworth, B. (1981). Genetics of life history in *Drosophila melanogaster*. I. Sib analysis of adult females. *Genetics.* 97, 173–186.

Rudman, SM, Greenblum, SI, Rajpurohit, S, Betancourt, NJ, Hanna, J, Tilk, S, … & Schmidt, P. (2021). Direct observation of adaptive tracking on ecological timescales in *Drosophila*. *bioRxiv*. doi.org/10.1101/2021.04.27.441526

Ruzicka F. (2018). The genetics and evolutionary dynamics of sexually antagonistic polymorphisms in *Drosophila melanogaster*. PhD thesis, University College London

Ruzicka, F, Hill, MS, Pennell, TM, Flis, I, Ingleby, FC, Mott, R, Fowler, K, Morrow, EH. & Reuter, M. (2019). Genome-wide sexually antagonistic variants reveal long-standing constraints on sexual dimorphism in fruit flies. *PLoS Biology*. 17(4), e3000244.

Ruzicka, F, Dutoit, L, Czuppon, P, Jordan, CY, Li, XY, Olito, C, … & Connallon, T. (2020). The search for sexually antagonistic genes: practical insights from studies of local adaptation and statistical genomics. *Evolution Letters*. 4(5), 398–415.

Ruzicka, F, Holman, L. & Connallon, T. (2021). Polygenic signals of sexually antagonistic selection in contemporary human genomes. *bioRxiv*. doi.org/10.1101/2021.09.20.461171

Ryding, S, Klaassen, M, Tattersall, GJ, Gardner, JL. & Symonds, MR. (2021). Shape-shifting: changing animal morphologies as a response to climatic warming. *Trends in Ecology & Evolution*. 36, 1036–1048.

Ryner, LC, Goodwin, SF, Castrillon, DH Anand, A, Villella, A, Baker, BS, … & Wasserman, SA. (1996). Control of male sexual behavior and sexual orientation in *Drosophila* by the fruitless gene. *Cell*. 87, 1079–1089.

Saminadin-Peter SS, Kemkemer C, Pavlidis P. & Parsch J. (2012). Selective sweep of a cis-regulatory sequence in a non-African population of *Drosophila melanogaster*. *Molecular Biology and Evolution*. 29, 1167–1174.

Sandler, G, Wright, SI. & Agrawal, AF. (2021). Patterns and Causes of Signed Linkage Disequilibria in Flies and Plants. *Molecular Biology and Evolution*. 38(10), 4310-4321.

dos Santos, G, Schroeder, AJ, Goodman, JL, Strelets, VB, Crosby, MA, Thurmond, J, … & FlyBase Consortium. (2015). FlyBase: introduction of the *Drosophila melanogaster* release 6 reference genome assembly and large-scale migration of genome annotations. *Nucleic Acids Research.* 43(D1), D690–D697.

Sato, DX. & Kawata, M. (2018). Positive and balancing selection on SLC18A1 gene associated with psychiatric disorders and human-unique personality traits. *Evolution Letters*. 2(5), 499–510.

Sato, K. & Yamamoto, D. (2020). The mode of action of Fruitless: Is it an easy matter to switch the sex? *Genes, Brain and Behavior*, 19(2), e12606.

Sato, MP, Makino, T. & Kawata, M. (2016). Natural selection in a population of *Drosophila melanogaster* explained by changes in gene expression caused by sequence variation in core promoter regions. *BMC Evolutionary Biology*. 16(1), 1–12.

Sayadi, A, Barrio, AM, Immonen, E, Dainat, J, Berger, D, Tellgren-Roth, C, … & Arnqvist, G., (2019). The genomic footprint of sexual conflict. *Nature ecology & evolution*. 3(12), 1725–1730.

Schäfer, D, Vincent, H, Fischer, M. & Kempel, A. (2020). The importance of genetic diversity for the translocation of eight threatened plant species into the wild. *Global Ecology and Conservation.* 24, e01240.

Schenkel, MA, Pen, I, Beukeboom, LW. & Billeter, JC. (2018). Making sense of intralocus and interlocus sexual conflict. *Ecology and Evolution.* 8, 13035–13050.

Schlötterer, C, Tobler, R, Kofler, R. & Nolte, V. (2014). Sequencing pools of individuals— mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics*. 15(11), 749–763.

Schlötterer, C, Kofler, R, Versace, E, Tobler, R. & Franssen, SU. (2015). Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation. *Heredity*. 114, 431–440.

Schou, MF, Loeschcke, V, Bechsgaard, J, Schlötterer, C. & Kristensen, TN. (2017). Unexpected high genetic diversity in small populations suggests

maintenance by associative overdominance. *Molecular Ecology*. 26(23), 6510–6523.

Schrider, DR. & Kern, A.D. (2018). Supervised machine learning for population genetics: a new paradigm. *Trends in Genetics*. 34(4), 301–312.

Seehausen, O, Terai, Y, Magalhaes, IS, Carleton, KL, Mrosso, HD, Miyagi, R, … & Imai, H. (2008). Speciation through sensory drive in cichlid fish. *Nature*. 455, 620–626.

Sellis, D, Callahan, BJ, Petrov, DA. & Messer, PW. (2011). Heterozygote advantage as a natural consequence of adaptation in diploids. *PNAS*, 108(51), 20666-20671.

Sham, P, Bader, JS, Craig, I, O'Donovan, M. & Owen, M. (2002). DNA pooling: a tool for large-scale association studies. *Nature Reviews Genetics*. 3, 862–871.

Siewert, KM. & Voight, BF. (2017). Detecting long-term balancing selection using allele frequency correlation. *Molecular Biology and Evolution*. 34, 2996–3005.

Sinervo B. & Lively CM. (1996). The rock–paper–scissors game and the evolution of alternative male strategies. *Nature.* 380, 240–243.

Sinervo, B, Miles, DB, Frankino, WA, Klukowski, M. & DeNardo, DF. (2000). Testosterone, endurance, and Darwinian fitness: natural and sexual selection on the physiological bases of alternative male behaviors in side-blotched lizards. *Hormones and Behavior*. 38(4), 222–233.

Sjödin PE, Sjöstrand A, Jakobsson M. & Blum MG. (2012). Resequencing data provide no evidence for a human bottleneck in Africa during the penultimate glacial period. *Molecular Biology and Evolution.* 29, 1851–1860.

Smith, DT, Hosken, DJ, Rostant, WG, Yeo, M, Griffin, RM, Bretman, A, … & Wedell, N. (2011). DDT resistance, epistasis and male fitness in flies. *Journal of Evolutionary Biology*. 24(6), 1351–1362.

Spencer, HG. & Priest, NK. (2016). The evolution of sex-specific dominance in response to sexually antagonistic selection. *The American Naturalist*. 187, 658–666.

Stearns, ASC. (1989). Trade-offs in life-history evolution. *Functional Ecology.* 3, 259–268.

Stephan, W. & Li, H. (2007). The recent demographic and adaptive history of *Drosophila melanogaster*. *Heredity*. 98(2), 65–68.

Stewart, KA, Draaijer, R, Kolasa, MR. & Smallegange, IM. (2019). The role of genetic diversity in the evolution and maintenance of environmentally-cued, male alternative reproductive tactics. *BMC Evolutionary Biology*. *19*, 1–10.

Svensson, EI, McAdam, AG. & Sinervo, B (2009). Intralocus sexual conflict over immune defense, gender load, and sex-specific signalling in a natural lizard population. *Evolution*. 63, 3124–3135.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics.* 123(3), 585–595.

Takahata, N. & Nei, M. (1990). Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics*. 124(4), 967–978.

Taus, T, Futschik, A. & Schlötterer, C. (2017). Quantifying selection with pool-seq time series data. *Molecular Biology and Evolution*. 34(11), 3023–3034.

Teixeira, JC. & Huber, CD. (2021). The inflated significance of neutral genetic diversity in conservation genetics. *PNAS.* 118(10). e2015096118

Tellier, A, Villaréal, LMMA. & Giraud, T. (2007). Antagonistic pleiotropy may help population-level selection in maintaining genetic polymorphism for transmission rate in a model phytopathogenic fungus. *Heredity.* 98, 45–52.

Therneau, TM. (2015). A Package for Survival Analysis in S. version 2.38.

Tilk, S, Bergland, A, Goodman, A, Schmidt, P, Petrov, D. & Greenblum, S. (2019). Accurate allele frequencies from ultra-low coverage pool-seq samples in evolve-and-resequence experiments. *G3: Genes, Genomes, Genetics.* 9(12), 4159–4168.

Tobler, R, Franssen, SU, Kofler, R, Orozco-terWengel, P, Nolte, V, Hermisson, J. & Schlötterer, C. (2014). Massive habitat-specific genomic response in *D. melanogaster* populations during experimental evolution in hot and cold environments. *Molecular Biology and Evolution.* 31(2), 364–375.

Trivers, RL. (1972). Parental investment and sexual selection. in B, Campbell. (ed) *Sexual Selection and the Descent of Man 1871–1971*. New York: Aldine de Gruyter, 136–179.

Turner, TL. & Miller, PM. (2012). Investigating natural variation in *Drosophila* courtship song by the evolve and resequence approach. *Genetics.* 191(2), 633–642.

Turner, TL, Stewart, AD, Fields, AT, Rice, WR. & Tarone, AM. (2011). Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in *Drosophila melanogaster*. *PLoS Genetics*. 7(3), e1001336.

Turner, TL, Miller, PM. & Cochrane, VA. (2013). Combining genome-wide methods to investigate the genetic complexity of courtship song variation in *Drosophila melanogaster*. *Molecular Biology and Evolution*. 30, 2113–2120.

Turelli, M. & Barton, NH. (2004). Polygenic variation maintained by balancing selection: pleiotropy, sex-dependant allelic effects and GxE interactions. *Genetics*. 166(2), 1053–1079.

Unckless, RL, Howick, VM. & Lazzaro, BP. (2016). Convergent balancing selection on an antimicrobial peptide in *Drosophila*. *Current Biology*. 26, 257–262.

Usui-Aoki, K, Ito, H, Ui-Tei, K, Takahashi, K, Lukacsovich, T, Awano, W, … & Yamamoto, D. (2000). Formation of the male-specific muscle in female *Drosophila* by ectopic fruitless expression. *Nature Cell Biology*. 2, 500–506.

Uyenoyama, MK. (2005). Evolution under tight linkage to mating type. *New Phytologist*. 165, 63–70.

Van der Auwera, GA, Carneiro, MO, Hartl, C, Poplin, R, Del Angel, G, Levy-Moonshine, A, … & Banks, E. (2013). From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*. 43(1), 11–10.

Van Doorn, GS. (2009). Intralocus sexual conflict. *Annals of the New York Academy of Sciences.* 1168, 52–71.

VanKuren, NW. & Long, M. (2018). Gene duplicates resolving sexual conflict rapidly evolved essential gametogenesis functions. *Nature Ecology and Evolution*. 2(4), 705–712.

Van Oosterhout, C, Joyce, DA, Cummings, SM, Blais, J, Barson, NJ, Ramnarine, IW, … & Cable, J. (2006). Balancing selection, random genetic drift, and genetic variation at the major histocompatibility complex in two wild populations of guppies (*Poecilia reticulata*). *Evolution*. 60, 2562–2574.

Van Valen, L. (1973) A new evolutionary law. *Evolutionary Theory.* 1, 1–30.

Vellend, M. & Geber, MA. (2005). Connections between species diversity and genetic diversity. *Ecology Letters*. 8(7), 767–781.

Vernes, SC. (2014). Genome wide identification of *Fruitless* targets suggests a role in upregulating genes important for neural circuit formation. *Scientific Reports*. 4, 4412.

Vicoso, B. & Charlesworth, B. (2009). Effective population size and the faster-X effect: an extended model. *Evolution*. 63, 2413–2426.

Waithe, D, Rennert, P, Brostow, G. & Piper, MD. (2015). QuantiFly: Robust trainable software for automated *Drosophila* egg counting. *PLoS ONE*. 10(5), e0127659.

Wardlaw, AM. & Agrawal, AF. (2019). Sexual conflict and sexually transmitted infections (STIs): coevolution of sexually antagonistic host traits with an STI. *The American Naturalist.* 193, E1–E14.

Wang, J, Wurm, Y, Nipitwattanaphon, M, Riba-Grognuz, O, Huang, YC, Shoemaker, D. & Keller, L. (2013). A Y-like social chromosome causes alternative colony organization in fire ants. *Nature*. 493, 664–668.

Wickham. H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

Wilkinson, GS, Breden, F, Mank, JE, Ritchie, MG, Higginson, AD, Radwan, J, … & Phillips, PC. (2015). The locus of sexual selection: moving sexual selection studies into the post-genomics era. *Journal of Evolutionary Biology*. 28, 739–755.

Williams GC. (1957). Pleiotropy, natural selection, and the evolution of senescence. *Evolution*. 11, 398–411.

Williams, PD. & Day, T. (2003). Antagonistic pleiotropy, mortality source interactions, and the evolutionary theory of senescence. *Evolution.* 57, 1478–1488.

Wittmann, MJ, Bergland, AO, Feldman, MW, Schmidt, PS. & Petrov, DA. (2017). Seasonally fluctuating selection can maintain polymorphism at many loci via segregation lift. *PNAS.* 114, E9932–E9941.

Wright, S. (1922). Coefficients of inbreeding and relationship. *The American Naturalist*, 56, 330–338.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*. *16*(2), 97–159.

Wright S. (1939). The distribution of self-sterility alleles in populations*. Genetics.* 24*,* 538–552.

Wright, S. (1968). *Evolution and the Genetics of Populations. Vol. I. Genetic and Biometric Foundations.* Chicago: Chicago University Press.

Xi, L, Schmidt, JC, Zaug, AJ, Ascarrunz, DR. & Cech, TR. (2015). A novel two-step genome editing strategy with CRISPR-Cas9 provides new insights into telomerase action and TERT gene expression. *Genome Biology.* 16, 1–17.

Yeaman, S. (2015). Local adaptation by alleles of small effect. *The American Naturalist*, 186, S74–S89.

Yeaman, S. & Otto, SP. (2011). Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution.* 65, 2123–2129.

Zajitschek, F. & Connallon, T. (2018). Antagonistic pleiotropy in species with separate sexes, and the maintenance of genetic variation in life-history traits and fitness. *Evolution.* 72, 1306–1316.

Zhu, Y, Chen, H, Fan, J, Wang, Y, Li, Y, Chen, J. … & Mew, TW. (2000). Genetic diversity and disease control in rice. *Nature*. 406, 718–722.

Appendix A

Creation of allelic lines

A.1 Creation of allelic lines

Isogenic allelic lines, which were fixed for either the S or L allele but otherwise isogenic for a Canton-S background were created in order to study the fitness effects of the *fru* polymorphism.

Hemiclonal LH_M flies (Rice et al. 2005) known to carry either the L or S allele were selected for the creation of the isogenic lines and were then backcrossed into a *Df(3R)fru*$^{4\text{-}40}$*/TM6B* stock. Flies of this stock carry chromosomes of an isogenic Canton-S genetic background, except for the third chromosome, where they are heterozygotes for a Canton-S chromosome carrying a deletion covering the *fru* locus (*Df(3R)fru*$^{4\text{-}40}$) (Anand et al. 2001), and the TM6B balancer chromosome. TM6B contains multiple and nested inversions and carries several homozygous lethal mutations, as well as dominant marker mutations which produce phenotypes for identification, including *Tubby* (*Tb*) that causes a distinct shape of the pupa (Miller et al. 2016).

Introgression of the *fru* allele onto the deficiency chromosome and into the Canton-S background was achieved by repeated rounds of backcrossing: (i) females heterozygous for a third chromosome carrying a focal *fru* allele (*fru*$^{S/L}$) and the *Df(3R)fru*$^{4\text{-}40}$ deficiency (themselves obtained by mating the hemiclonal line and females from the *Df(3R)fru*$^{4\text{-}40}$ deficiency stock), with (ii) males from *Df(3R)fru*$^{4\text{-}40}$ deficiency stock (see Figure A.1). Since balancer and deficiency chromosomes are lethal in homozygous state and balancers carry the dominant *Tb* marker, the wild-type offspring of a hemiclone/*Df(3R)fru*$^{4\text{-}40}$ x *Df(3R)fru*$^{4\text{-}40}$/TM6B cross are always identifiable as *fru*$^{S/L}$/*Df(3R)fru*$^{4\text{-}40}$ heterozygotes. By repeatedly backcrossing *fru*$^{S/L}$/*Df(3R)fru*$^{4\text{-}40}$ heterozygote females to *Df(3R)fru*$^{4\text{-}40}$/TM6B males, the original hemiclonal genome carrying the focal *fru* allele is gradually eroded through recombination in females and replaced with the isogenic Canton-S background of the *Df(3R)fru*$^{4\text{-}40}$ deficiency line. After 7 generations of

backcrossing, the allelic lines should carry on average less than 1% of the original hemiclonal haplotype (Abbott and Morrow 2011).

Having introgressed the *fru* allele into the Canton-S background of *Df(3R)fru*$^{4\text{-}40}$, lines homozygous for the *fru* allele were created (as opposed to *fru*$^{S/L}$/*Df(3R)fru*$^{4\text{-}40}$ heterozygotes). Because *fru*$^{S/L}$/*Df(3R)fru*$^{4\text{-}40}$ heterozygotes and *fru*$^{S/L}$/ *fru*$^{S/L}$ homozygotes are phenotypically indistinguishable, this was achieved through a two-step crossing procedure. An initial cross served to identify pairs of parents in which both individuals carried a focal *fru* allele. Virgin *Tb*-carrying offspring of a *fru*$^{S/L}$/*Df(3R)fru*$^{4\text{-}40}$ x *Df(3R)fru*$^{4\text{-}40}$/TM6B cross (either *fru*$^{S/L}$/TM6B or *Df(3R)fru*$^{4\text{-}40}$/TM6B) were set up in pairs (dyads A, B, C, see “Phase 3” in Figure A.1). Depending on the genotypes of the F1 pair, this cross can either produce: (i) 100% *Tb* F2s, if both F1 parents were *Df(3R)fru*$^{4\text{-}40}$/TM6B—these were discarded, or (ii) some fraction of non-*Tb* F2s, if the F1 pair were *fru*$^{S/L}$/TM6B+*Df(3R)fru*$^{4\text{-}40}$/TM6B or *fru*$^{S/L}$/TM6B+*fru*$^{S/L}$/TM6B. To distinguish the two latter cases and identify pairs of *fru*$^{S/L}$/TM6B individuals that are capable of producing the *fru*$^{S/L}$/ *fru*$^{S/L}$ individuals required, an additional ‘test cross’ was performed where F2s were backcrossed to *Df(3R)fru*$^{4\text{-}40}$/TM6 males. Based on the F3 phenotype, the genotype of the F2 could be inferred, as *fru*$^{S/L}$/ *fru*$^{SL/}$ F2s produce a 1:1 ratio of wild-type to *Tb* F3s, whereas *fru*$^{S/L}$/*Df(3R)fru*$^{4\text{-}40}$ heterozygotes produce 1:2 ratio of wild-type to *Tb* F3s. F2s producing a ratio of wild-type to *Tb* F3s that was significantly less than 1:2 (as assessed from a χ^2 test) were used to establish isogenic allelic lines. This approach generated three independent isogenic lines each for the S and L allele. The full crossing scheme is shown in Figure A.1.

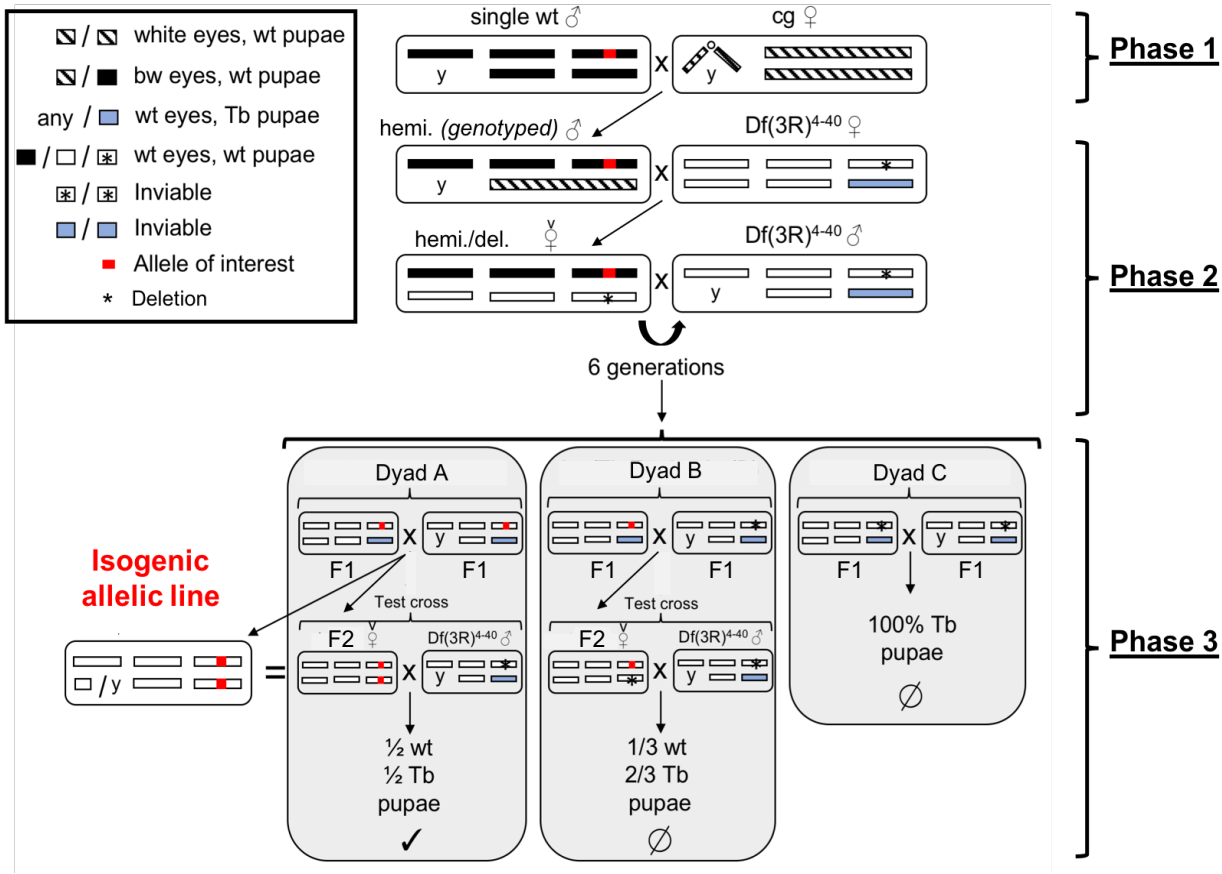


Figure A1. Crossing scheme used to create isogenic lines. Wildtype (wt) flies are cross with the clone generator (cg) stock. The offspring were then crossed to a *fru* deletion stock *Df(3R)fru*$^{4\text{-}40}$ female. The result of this cross was back crossed with males of the *Df(3R)fru*$^{4\text{-}40}$ stock for seven generations to complete the introgression of the *fru* allele. The resulting flies have wildtype pupa and can be distinguished from non-isogenic lines which have the Tb pupa phenotype. Clone generator line flies carry compound chromosomes preventing recombination (Rice 1996). Tb – flies display the 'tubby' pupal phenotype.

Appendix B

A non-coding indel polymorphism in the *fruitless* gene of *Drosophila melanogaster* exhibits antagonistically pleiotropic fitness effects

PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Research

Cite this article: Jardine MD, Ruzicka F, Diffley C, Fowler K, Reuter M. 2021 A non-coding indel polymorphism in the *fruitless* gene of *Drosophila melanogaster* exhibits antagonistically pleiotropic fitness effects. *Proc. R. Soc. B* **288**: 20202958. https://doi.org/10.1098/rspb.2020.2958

Received: 26 November 2020
Accepted: 14 April 2021

Subject Category:
Evolution

Subject Areas:
evolution

Keywords:
balancing selection, antagonistic pleiotropy, *fruitless*, *Drosophila*, genetic variation

Author for correspondence:
Michael D. Jardine
e-mail: michael.jardine.17@ucl.ac.uk

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.5410945.

THE ROYAL SOCIETY PUBLISHING

A non-coding indel polymorphism in the *fruitless* gene of *Drosophila melanogaster* exhibits antagonistically pleiotropic fitness effects

Michael D. Jardine[1,2], Filip Ruzicka[3], Charlotte Diffley[1], Kevin Fowler[1,2] and Max Reuter[1,2]

[1]Department of Genetics, Evolution and Environment, University College London, London, UK
[2]Centre for Life's Origins and Evolution, University College London, London, UK
[3]School of Biological Sciences and Centre for Geometric Biology, Monash University, Clayton, Australia

MDJ, 0000-0001-6407-105X; FR, 0000-0001-9089-624X; KF, 0000-0001-9737-7549; MR, 0000-0001-9554-0795

The amount of genetic variation for fitness within populations tends to exceed that expected under mutation–selection–drift balance. Several mechanisms have been proposed to actively maintain polymorphism and account for this discrepancy, including antagonistic pleiotropy (AP), where allelic variants have opposing effects on different components of fitness. Here, we identify a non-coding indel polymorphism in the *fruitless* gene of *Drosophila melanogaster* and measure survival and reproductive components of fitness in males and females of replicate lines carrying each respective allele. Expressing the *fruitless* region in a hemizygous state reveals a pattern of AP, with one allele generating greater reproductive fitness and the other conferring greater survival to adulthood. Different fitness effects were observed in an alternative genetic background, which may reflect dominance reversal and/or epistasis. Our findings link sequence-level variation at a single locus with complex effects on a range of fitness components, thus helping to explain the maintenance of genetic variation for fitness. Transcription factors, such as *fruitless*, may be prime candidates for targets of balancing selection since they interact with multiple target loci and their associated phenotypic effects.

Electronic supplementary material for:

A non-coding indel polymorphism in the *fruitless* gene of *Drosophila melanogaster* exhibits antagonistically pleiotropic fitness effects

Michael D. Jardine[1,2], Filip Ruzicka[3], Charlotte Diffley[1], Kevin Fowler[1,2], Max Reuter[1,2]

[1]Department of Genetics, Evolution and Environment, University College London, London, United Kingdom
[2]Centre for Life's Origins and Evolution, University College London, London, United Kingdom
[3]School of Biological Sciences and Centre for Geometric Biology, Monash University, Clayton, Australia

DOI: 10.1098/rspb.2020.2958

Methods S1 – Identification of a polymorphic indel in the *fruitless* gene

Signatures of balancing selection along the *fru* gene

We investigated signatures of balancing selection along the *fru* gene in two wild population samples of *D. melanogaster* flies: a North American population sample of 205 genomes (RAL) and a Zambian population sample of 197 genomes (ZI) [1,2]. Elevated polymorphism and linkage disequilibrium (LD) can both indicate that a given region is under balancing selection [3]. We therefore estimated regional polymorphism (nucleotide diversity, Tajima's D) and regional LD (Kelly's ZnS) over 1000bp windows (500bp step) along the *D. melanogaster* (release 6) genome, in each population, using PopGenome [4].

Sanger sequencing of a candidate *fru* region

A ~1000bp region of the *fru* gene was identified as exhibiting elevated levels of polymorphism and LD in both North American and Zambian population samples (Figure 1, main text; Results S1). To investigate this region in more detail, 96 chromosomes were sampled from LH_M, a laboratory-adapted North American population of *D. melanogaster* [5]. Sampling was performed using a 'hemiclonal' approach, in which purpose-built 'clone generator' flies are used to manipulate haploid chromosome sets (X, II, III) [6]. Individual hemiclonal males were crossed with females from a deficiency strain (*Df(3R)BSC509*), which carries a deletion spanning the *fru* gene and a TM6C balancer complement marked with *Stubble* (*Sb*). DNA from the hemiclone/*Df(3R)BSC509* heterozygote offspring of this cross was extracted using standard protocols (see "Phase 1" in Figure S1). A ~400bp region of the *fru* gene was then PCR-amplified and Sanger-sequenced using the following primers: 5'-CACCCAACGCCACCTAGTTA-3' (forward) and 5'-CGCCACTTGATTGCCACATT-3' (reverse).

Balancer stock genotyping

To ascertain the *fru* allele carried by the TM6B balancer, DNA was extracted from several *Df(3R)fru$^{4\text{-}40}$/TM6B* flies and the indel region was then PCR-amplified as above. The size of the PCR product was checked on an agarose gel (using control reaction with L- and S-bearing DNA templates as controls) and Sanger-sequenced to confirm allelic identity.

Results S1 - Identification of a polymorphic indel in the *fruitless* gene

We found that a 1000bp-window of the *fru* gene exhibited unusually high levels of polymorphism and local LD relative to the genome-wide average (red dashed line in Figure 1, main text). This was true both in the RAL population (upper 2nd percentile of nucleotide diversity; upper 12th percentile of Tajima's D; upper 5th percentile of Kelly's ZnS), and in the ZI population (upper 5th percentile of nucleotide diversity; upper 11th percentile of Tajima's D; upper 9th percentile of Kelly's ZnS). Sanger

sequencing further revealed that this polymorphic region of *fru* segregates for a 43bp indel, producing fragment length differences between the PCR products of the two alternative haplotypes in this region. We therefore designated these haplotypes 'Long' (L) and 'Short' (S), respectively.

To infer the frequency of the *fru* indel polymorphism in the RAL and ZI populations in the absence of direct indel polymorphism data, we examined the frequency of SNPs located in very close proximity to (<80bp) and in tight LD (in LH_M) with the indel (Figure 1, main text). A haplotype network constructed from these SNPs showed that haplotypes do not cluster by population but fall into divergent allelic classes that occur at intermediate frequencies in both populations (Figure 1, main text). Given the large evolutionary distances between the RAL and ZI populations used in the construction of the haplotype network, this is suggestive evidence that the *fru* indel (and/or alleles linked to it) are under some form of antagonistic and/or balancing selection. We therefore performed further experiments to test this hypothesis.

Methods S2 – Creation of isogenic lines

To assess the sex-specific fitness effects of the L and S alleles, we created fly lines homozygous for each allele but otherwise isogenic for a Canton-S background across the rest of their genome ('isogenic allelic lines'; see Figure S1 for the full crossing scheme).

First, we randomly selected three lines carrying the S allele and three lines carrying the L allele among the 96 sequenced hemiclonal lines (see Methods S1, "Sanger sequencing of a candidate *fru* region") and introgressed these alleles into an isogenic background, as described below. Introgression of the *fru* allele was performed with the help of a *Df(3R)fru*$^{4\text{-}40}$/TM6B deficiency stock, carrying a deletion spanning the *fru* locus (see Figure S1) in a Canton-S background,

complemented with the third-chromosome balancer TM6B marked with the dominant mutation *Tubby* (*Tb*). Introgression of the *fru* allele onto the deficiency chromosome and into the Canton-S background was achieved by repeatedly backcrossing: (i) females heterozygous for a third chromosome carrying a focal *fru* allele ($fru^{S/L}$) and the $Df(3R)fru^{4-40}$ deficiency (themselves obtained by mating the hemiclonal line and females from the $Df(3R)fru^{4-40}$ deficiency stock), with (ii) males from $Df(3R)fru^{4-40}$ deficiency stock (see Figure S1). Since balancer and deficiency chromosomes are lethal in homozygous state and balancers carry the dominant *Tb* marker, the wild-type offspring of a hemiclone/$Df(3R)fru^{4-40}$ x $Df(3R)fru^{4-40}$/TM6B cross are always identifiable as $fru^{S/L}$/$Df(3R)fru^{4-40}$ heterozygotes. By repeatedly backcrossing $fru^{S/L}$/$Df(3R)fru^{4-40}$ heterozygote females to $Df(3R)fru^{4-40}$/TM6B males, the original hemiclonal genome carrying the focal *fru* allele is gradually eroded through recombination in females and replaced with the isogenic Canton-S background of the $Df(3R)fru^{4-40}$ deficiency line. After 7 generations of backcrossing, the allelic lines should carry on average less than 1% of the original hemiclonal haplotype (i.e. 1% of the original X-II-III complement).

Having introgressed the *fru* allele into the Canton-S background of $Df(3R)fru^{4-40}$, we created lines homozygous for the *fru* allele (as opposed to $fru^{S/L}$/$Df(3R)fru^{4-40}$ heterozygotes). Because $fru^{S/L}$/$Df(3R)fru^{4-40}$ heterozygotes and $fru^{S/L}$/ $fru^{S/L}$ homozygotes are phenotypically indistinguishable, this was achieved through a two-step crossing procedure. An initial cross served to identify pairs of parents in which both individuals carried a focal *fru* allele. Virgin *Tb*-carrying offspring of a $fru^{S/L}$/$Df(3R)fru^{4-40}$ x $Df(3R)fru^{4-40}$/TM6B cross (either $fru^{S/L}$/TM6B or $Df(3R)fru^{4-40}$/TM6B) were set up in pairs (dyads A, B, C, see "Phase 3" in Figure S1). Depending on the genotypes of the F1 pair, this cross can either produce: (i) 100% *Tb* F2s, if both F1 parents were $Df(3R)fru^{4-40}$/TM6B—these were discarded, or (ii) some fraction of non-*Tb* F2s, if the F1 pair were $fru^{S/L}$/TM6B+$Df(3R)fru^{4-40}$/TM6B or $fru^{S/L}$/TM6B+$fru^{S/L}$/TM6B. To distinguish the two latter cases and identify pairs of $fru^{S/L}$/TM6B individuals that are capable of producing the $fru^{S/L}$/ $fru^{S/L}$ individuals we

required, an additional 'test cross' was performed where F2s were backcrossed to *Df(3R)fru*$^{4\text{-}40}$/TM6 males. Based on the F3 phenotype, the genotype of the F2 could be inferred, as *fru*$^{S/L}$/ *fru*$^{SL/}$ F2s produce a 1:1 ratio of wild-type to *Tb* F3s, whereas *fru*$^{S/L}$/*Df(3R)fru*$^{4\text{-}40}$ heterozygotes produce 1:2 ratio of wild-type to *Tb* F3s. F2s producing a ratio of wild-type to *Tb* F3s that was significantly less than 1:2 (as assessed from a χ^2 test) were used to establish isogenic allelic lines.

References

1. Mackay TFC *et al.* 2012 The *Drosophila melanogaster* Genetic Reference Panel. *Nature* **482**, 173–178. (doi:10.1038/nature10811)
2. Lack JB, Cardeno CM, Crepeau MW, Taylor W, Corbett-Detig RB, Stevens KA, Langley CH, Pool JE. 2015 The *Drosophila* Genome Nexus: A population genomic resource of 623 *Drosophila melanogaster* genomes, including 197 from a single ancestral range population. *Genetics* **199**, 1229–1241. (doi:10.1534/genetics.115.174664)
3. Fijarczyk A, Babik W. 2015 Detecting balancing selection in genomes: Limits and prospects. *Mol. Ecol.* **24**, 3529–3545. (doi:10.1111/mec.13226)
4. Pfeifer B, Wittelsburger U, Ramos-Onsins SE, Lercher MJ. 2014 PopGenome: An efficient swiss army knife for population genomic analyses in R. *Mol. Biol. Evol.* **31**, 1929–1936. (doi:10.1093/molbev/msu136)
5. Rice WR, Linder JE, Friberg U, Lew TA, Morrow EH, Stewart AD. 2005 Inter-locus antagonistic coevolution as an engine of speciation: assessment with hemiclonal analysis. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 6527–6534. (doi:10.1073/pnas.0501889102)
6. Abbott JK, Morrow EH. 2011 Obtaining snapshots of genetic variation using hemiclonal analysis. *Trends Ecol. Evol.* **26**, 359–368. (doi:10.1016/j.tree.2011.03.011)

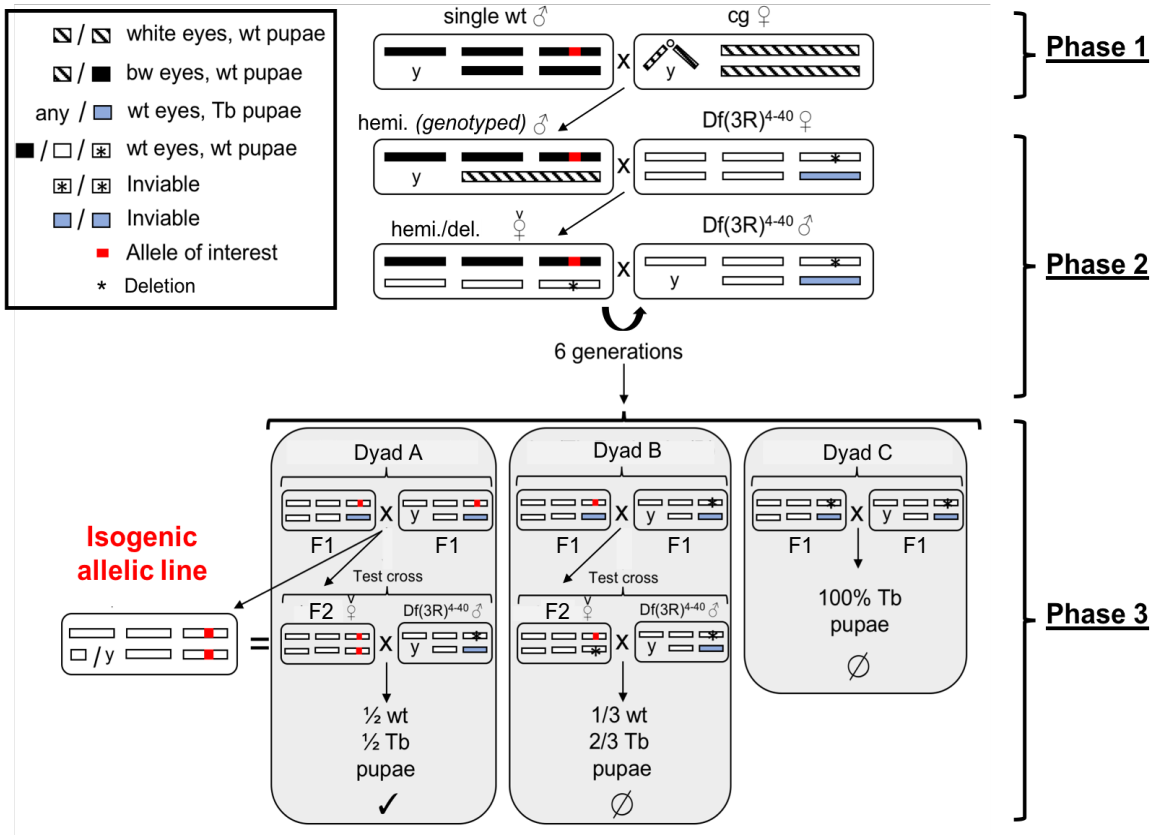


Figure S1. Crossing scheme used to create isogenic lines. See Methods S2 for details.

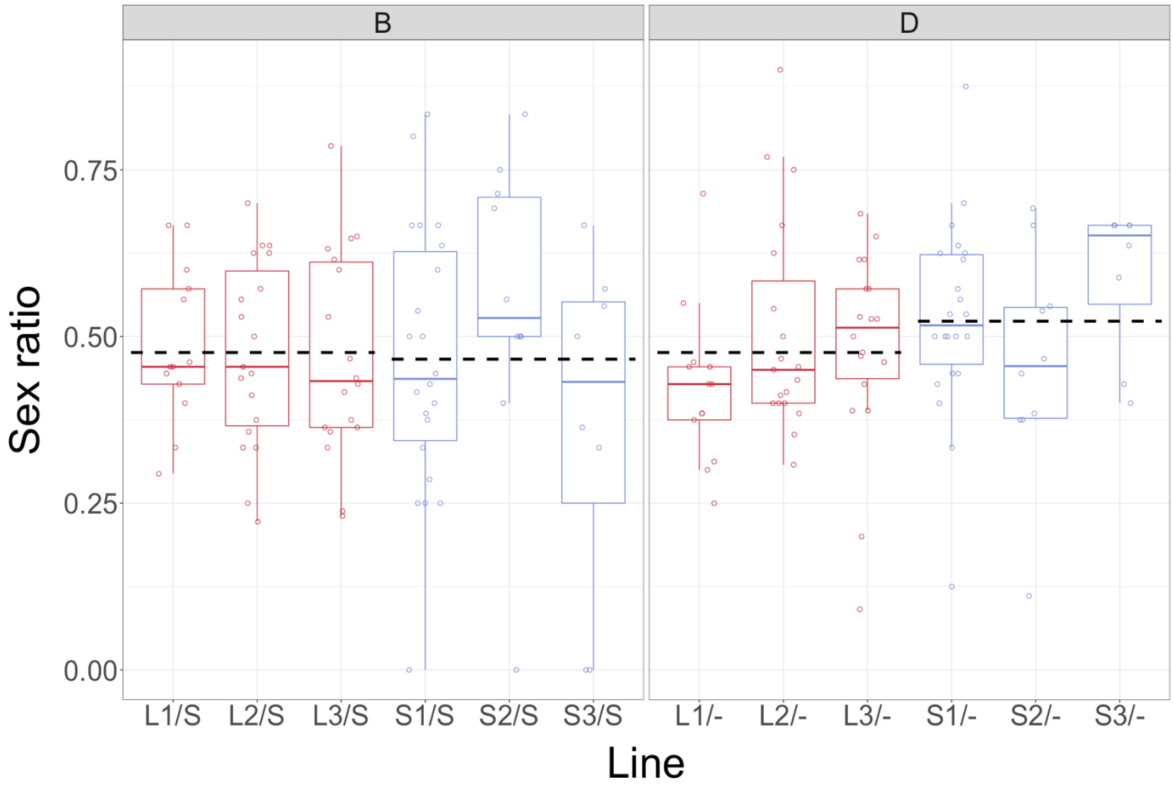


Figure S2. Sex ratio among surviving offspring presented for each line (L1-3 and S1-3) and chromosomal complement (B and D). Allelic means represented by dashed lines (L/B: 0.476±0.019; S/B: 0.466±0.035; L/D: 0.477±0.021; S/D: 0.0523±0.024). Sex ratio is defined as the proportion of males among offspring at eclosion.

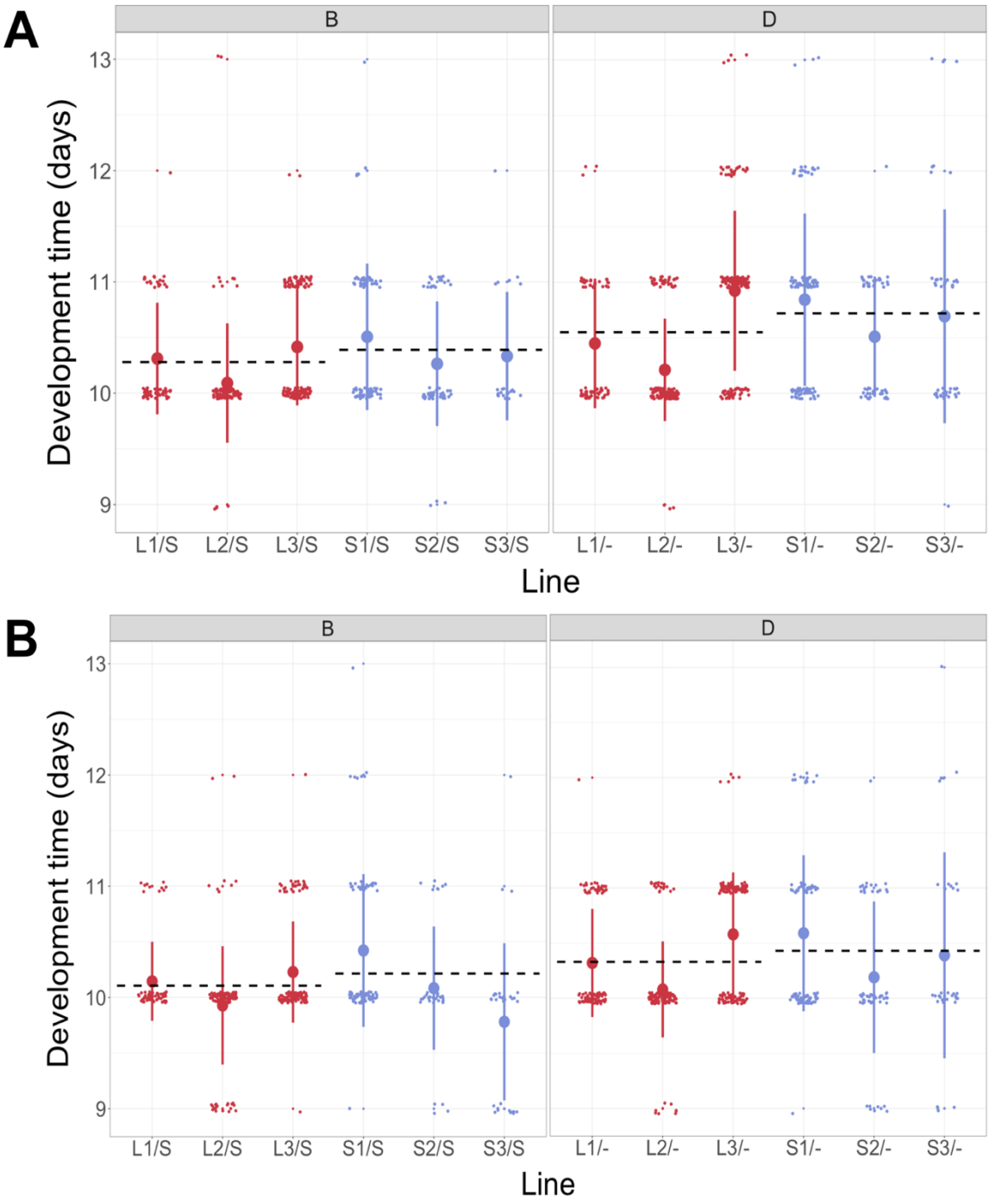

A
B
D
Development time (days)
13
12
11
10
9
L1/S
L2/S
L3/S
S1/S
S2/S
S3/S
L1/-
L2/-
L3/-
S1/-
S2/-
S3/-
Line
B
D
Development time (days)
13
12
11
10
9
L1/S
L2/S
L3/S
S1/S
S2/S
S3/S
L1/-
L2/-
L3/-
S1/-
S2/-
S3/-
Line

Figure S3. Development time (days ±standard error) of *fru* allelic lines (L1-3 and S1-3), for each chromosomal complement (B and D). Allelic means represented by dashed lines. Since sex was the most important factor in determining development time, this data is presented with the sexes separated: **A)** male flies (L/B: 10.28±0.03; S/B: 10.4±0.05; L/D: 10.55±0.036; S/D: 10.72±0.054), and **B)** female flies (L/B: 10.1±0.027; S/B: 10.22±0.056; L/D: 10.33±0.028; S/D: 10.42±0.056).

Table S1. Results from Cox Proportional Hazard (CPH) models applied to lifespan data. Five models were used. One was for all flies and then the data was split to have separate models for each chromosome complement (B and D) and sex (female or male). The first column indicates the set of data the model is applied to, while the second column indicates the term being tested in that model. CPH models use one level of a term as the reference level with a value of one. Other levels are then compared to this. The comparison made is shown in brackets as: (compared level:reference). Each term in a model has a hazard-ratio (H-R), a 95% confidence interval and a H-R p-value, which indicates if the compared level differs from the reference level. Also presented are χ_1^2 and its p-value, indicating the contribution of each term to the overall risk of mortality.

Model	Term (comparison)	HR	95%-CI	HR p-value	χ_1^2	p-value
All flies	*fru* allele (S:L)	1.318	1.126-1.544	<0.001	0.139	0.71
	Complement (D:B)	0.519	0.44-0.612	<0.001	43.79	<0.001
	Sex (M:F)	0.531	0.449-0.627	<0.001	31.886	<0.001
	Allele x complement (S/D:L/F)	0.693	0.57-0.841	<0.001	10.411	0.0013
	Allele x sex (S/D:L/B)	0.821	0.676-0.997	0.046	4.856	0.0276
	Complement x sex (D/M:B/F)	2.624	2.154-3.198	<0.001	90.752	<0.0001
	Allele x complement x sex (S/D/M:L/B/F)	1.258	0.852-1.856	0.249	1.331	0.249
B only	*fru* allele (S:L)	1.386	1.16-1.655	<0.001	3.848	0.049
	Sex (M:F)	0.572	0.472-0.692	<0.001	105.65	<0.001
	Allele x sex (S/D:L/B)	0.731	0.561-0.953	0.02	5.368	0.021
D only	*fru* allele (S:L)	0.87	0.715-1.059	0.164	5.317	0.021
	Sex (M:F)	1.32	1.081-1.614	0.0066	10.705	0.001
	Allele x sex (S/D:L/B)	0.927	0.696-1.234	0.604	0.269	0.604
Females only	*fru* allele (S:L)	1.381	1.157-1.65	<0.001	2.334	0.127
	Complement (D:B)	0.542	0.449-0.655	<0.001	14.879	<0.001
	Allele x complement (S/D:L/F)	0.611	0.469-0.798	<0.001	13.127	<0.0001
Males only	*fru* allele (S:L)	1.039	0.854-1.263	0.705	1.276	0.259
	Complement (D:B)	1.301	1.061-1.595	0.011	3.119	0.077
	Allele x complement (S/D:L/F)	0.772	0.58-1.029	0.077	3.117	0.077

Appendix C

Development of summary statistics and selection models

C.1 Introduction

The following sections relate to Chapter 3 and provide additional methods and results regarding the development of the selection models, and the relationship between the summary statistics and the different model classes, all of which informed the methodology described in Chapter 3. Additional cross-validation of the ABC parameter estimation is also included.

C.2 Methods

To diagnose the mode of selection acting upon the *fruitless* (*fru*) polymorphism by approximate Bayesian computation (ABC), five selection models were investigated: *neutrality*, *positive selection*, *negative selection*, *positive balancing selection*, and *negative balancing selection*. These models simulated selection under these different regimes in the 10 cage populations used in the experiment described in Chapter 3.

Simulations were performed in SLiM 3 (Haller and Messer, 2019), using custom scripts. These scripts were composed of a standard Wright-Fisher model of a single locus with two alleles (the L and S alleles). The simulations were designed to mimic the structure of the actual experiment as closely as possible. Simulations included ten subpopulations representing the ten cages. There was no migration between these subpopulations. Five subpopulations started with a frequency of 0.9 and the other five with a frequency of 0.1 for the measured allele to match the starting frequencies of the *fru* polymorphism. The measured allele is the one that SLiM tracks and reports the frequencies of. This is the equivalent of the S allele in the real data for easier comparison with the proxy SNP allele trajectories. All individuals at the start of the simulation were homozygotes at the *fru* locus as in the actual experiment. A generation time of two generations per month was assumed. Each simulation was run for 56 generations to correspond to the final sampled time point 28 months after the start of the experiment. Effective population size, *Ne,* was set at 1000, which is a very conservative

estimate. The relative fitness of the two alleles was determined by two parameters: the selection coefficient, *s*, and the dominance coefficient, *h*. The values of these parameters are chosen from a uniform distribution within a parameter range.

For neutrality, both parameters always have a value of 0 since no selection occurs. Two directional selection models, *positive* and *negative selection,* are similar but differ in that the focal allele is fitter (positive) or less fit (negative) than the alternative allele. For *positive selection*, *s* is drawn from between 0 and 1, while for *negative selection*, *s* is drawn from between -1 and 0. For both, *h* is between 0 and 1. Balancing selection was modelled using a heterosis model where the heterozygote is fitter than either homozygote. Although this is not the process of balancing selection hypothesised to be acting at the *fru* locus (Chapter 2), heterosis is a far easier method by which to model balancing selection by altering *h* to be >1. This still allows the likelihood of balancing selection acting at this locus to be studied. The different balancing selection models, *positive* and *negative balancing selection*, are distinguished based on which homozygote, SS or LL, is the fitter of the two, thus shifting the theoretical balancing point of the polymorphism. As with *directional selection*, *s* was drawn from a range of between 0 and 1 for positive and between -1 and 0 for *negative balancing selection*. Dominance, *h*, ranged from between 1 and 10.

One million simulation runs were performed for each selection model. The frequency of the focal allele in each population at the end of generations 2, 4, 8, 12, 20, 28, 36, 44 and 56, was recorded in line with the sampling times of the cages. Frequencies were organised into a 10x9 matrix for each simulation run, with one row per cage, with each column the nine sampling points ordered by collection date. Frequencies were then subjected to two rounds of binomial sampling to mirror the two sampling stages of the experiment. The first was the selection of 96 chromosomes (48 flies) from each population to represent the cage sampling. The second represented the

number of reads (coverage depth) covering each locus. This second step was repeated separately for each proxy using the coverage values in each pool. This process resulted in adding more noise to the simulated data, which should then be more comparable to the real sequencing results.

Refinement of simulations

13 selection summary statistics were formulated to summarise and quantify the frequency trajectories. Each summary statistic was designed to help distinguish between different selection models, for example SNPs under directional selection will more likely fix than those under balancing selection. The usefulness of each summary statistic was tested using correlation matrices and violin plots to ensure that each metric contributed some information in distinguishing later selection models. Summary statistics were formulated to describe an aspect of the whole experiment where a single number represents something about all 10 cages.

The 13 selection summary statistics were then calculated for each simulation run in the same manner as for each proxy SNP, creating a large dataset of simulated summary statistic values to compare with the real data. I created a series of histograms, one of each summary statistic per model, to visualise the spread of simulated summary statistic values compared to the real ones. This was to see how the simulations matched our data, whether they were capturing the variation they were designed to, and if they would be useful in the ABC analysis later. I then tested the proportion of overlap between the SNP values and the simulation values by calculating the proportion of simulations where their metric value was greater than the SNP value for each summary statistic and model. For summary statistic 2, persistence time, this was changed to calculate the proportion of simulated values smaller than the SNP value since the maximum value possible was 56 generations, the length of the experiment. These values were compiled into a table to observe the number of summary statistics where there was reasonable overlap between the simulated and real values. Reasonable

overlap was defined for summary statistics 3-13 as when the proportion of simulated metrics reporting a higher value than the SNP was between 0.05 and 0.95. Since summary statistics 1 and 2 have truncated non-continuous distributions, reasonable overlap was defined as being where <5% of simulations had a larger (summary statistic 1) or lower (summary statistic 2) value than the SNP value. Any model which produced summary statistic values with <5% overlap were considered unlikely to have produced the patterns observed and the proxy SNP summary statistics.

To observe how the summary statistics changed with the selection parameters, I also plotted a random subset of 1000 simulated metric values against the *s* and *h* values used in those simulations. By plotting the SNP value over these we could observe the approximate range of parameter values that produced simulated metric values similar to the SNP values. In doing this I noticed that some values of *s* and h never produced metric values remotely close to those calculated for the three proxy SNPs. Reducing the simulation set to only those where *s* was between 0 and 0.25 for positive selection, or -0.25 and 0 for *negative selection* showed a much better fit. I therefore decided to repeat the simulation process using simulations with a refined range of parameter values.

Refinement of simulations

The selection coefficient, *s,* used in the simulations was reduced to a maximum value of 0.25 for positive selection and a minimum of -0.25 for negative selection. This was repeated for the equivalent balancing selection models. For balancing selection, dominance was reduced to a range of between 1 and 5. The ranges of the refined selection parameters are summarised in Table 3.2. One million simulation runs were produced for each model, but this time using the refined range of *s* and *h* values. Selection metrics were calculated for these simulations with histograms and overlap values produced in the same way as before. I also created a series of violin plots to help visualise the distribution of the various summary statistics and to

ensure that the patterns for each metric were different between selection models. This would help the ABC analysis to discern between the various selection models later.

C.3 Results

In the following presentation of the results, for brevity, I sometimes only display figures for SNP 1521, which has the highest sequencing cover and therefore presents the most accurate results. But I note that differences between the SNPs were small and don't alter the overall conclusions.

Values of the proportion of overlap between simulated summary statistics from the five selection models and the proxy SNP summary statistics are presented in Table C1. There was little overlap in 8 out of 13 metrics for neutrality, with some metrics showing zero overlap between simulated and SNP values, Table C1. There were inconsistent patterns in the number of summary statistics displaying reasonable overlap between simulations and proxy SNPs for all other models and SNPs, but all had more summary statistics with at least some overlap than the *neutrality* model. Plots comparing selection metrics *s* and *h* parameter values with the metrics they produced are shown in Figure C1. These show that there was little variation in metric values once *s* exceeds 0.25 for positive selection, or once *s* is <-0.25 for negative selection, due to rapid fixation of the allele in all cages. This led to the decision to refine the range of *s* and *h* parameter values used in this study.

Refinement of selection parameters was performed and the proportion of overlap between the simulated summary statistics from the five selection models and the three proxy SNPs calculated as before. These are shown in Table C2 for SNP1521. For SNP 1521, refined parameter *positive balancing selection* displayed reasonable overlap with 11/13 summary statistics, compared to 9/13 for both refined parameter *positive* and *negative directional selection*, and 10/13 for refined parameter *negative balancing selection*,

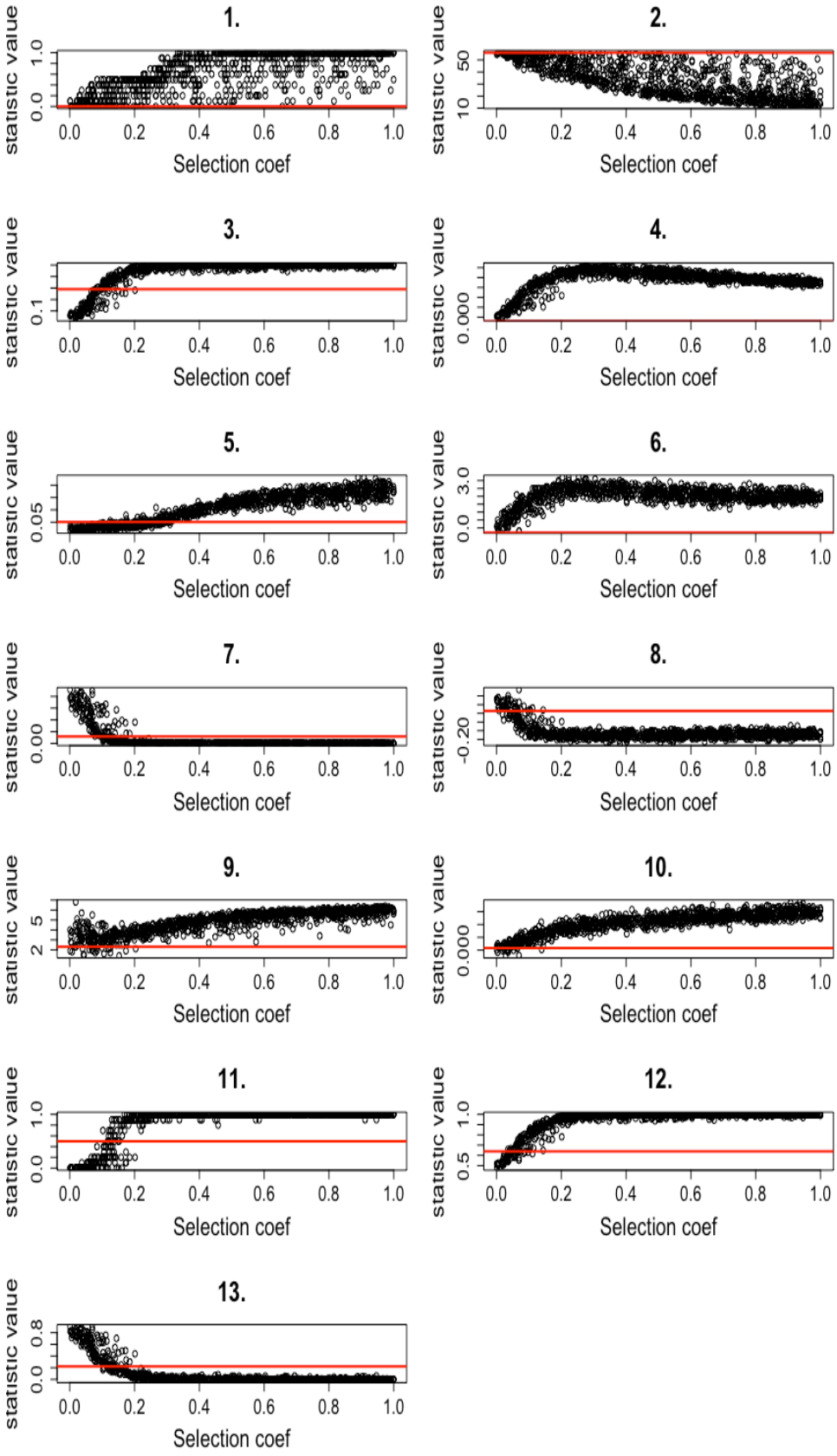
Table C2. There was no overlap with metric 12 for *negative directional selection* or *negative balancing selection* models with refined parameters. Overall, the refinement of selection parameters led to a greater degree of overlap between simulations summary statistics and proxy SNP summary statistics. So the simulations are capturing a better range of values, closer to those occurring in the experiment. Therefore this refined range of parameter values were used to create the models used for model choice and parameter estimation using ABC in Chapter 3.

Histograms of the simulated summary statistics were produced for all models, both pre and post parameter refinement. Examples are shown in Figure C2. Violin plots showing the range summary statistics for each refined model and neutrality are shown in Figure C3. These plots helped ensure that each summary statistic added some degree of information to the ABC analysis as I evaluated statistics and rejected those that did not show different patterns between the selection models.

ABC parameter estimation cross validation

Cross validation was required to see which method and tolerance rate was optimal for estimating the parameter values for our data. These results are presented in Table C3 and Figure C4. These show that the neural net method produced the smallest estimated error rate for estimating both *s*, and *h,* in the three SNPs.

A



B

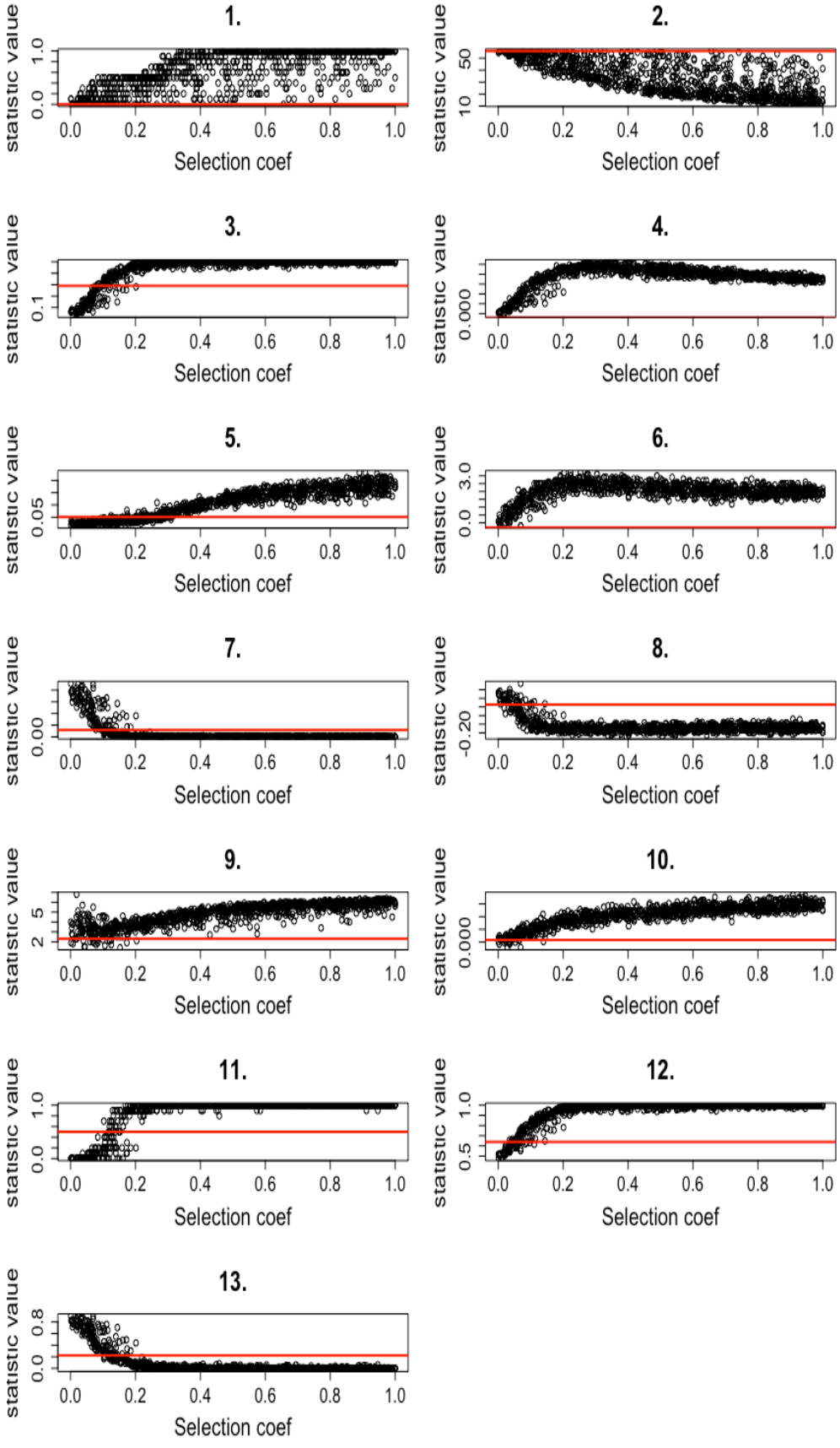
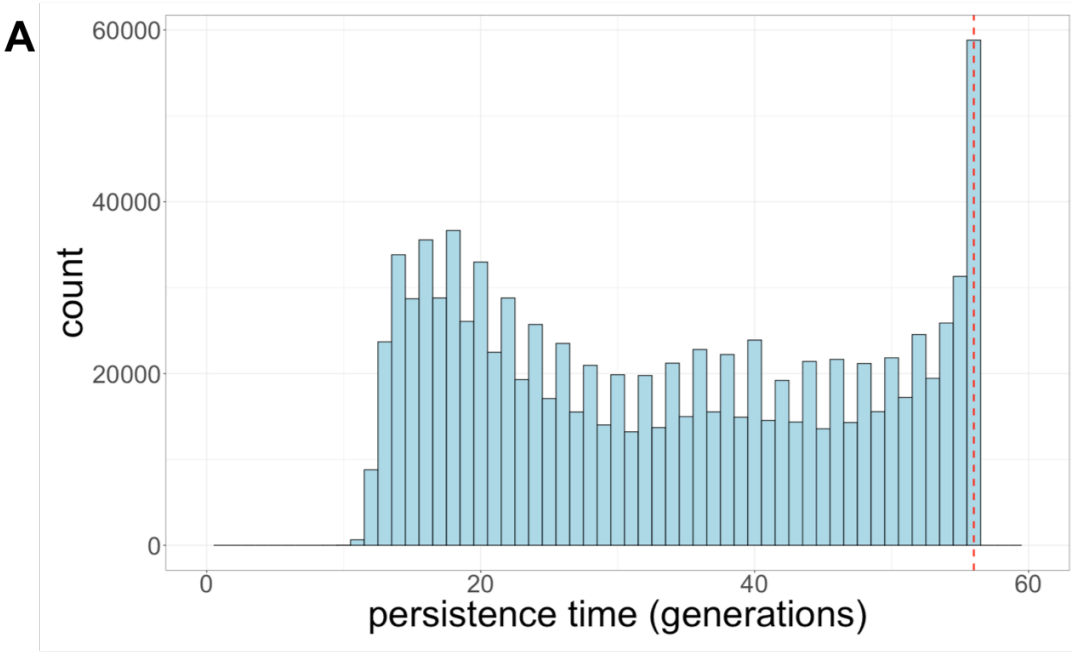
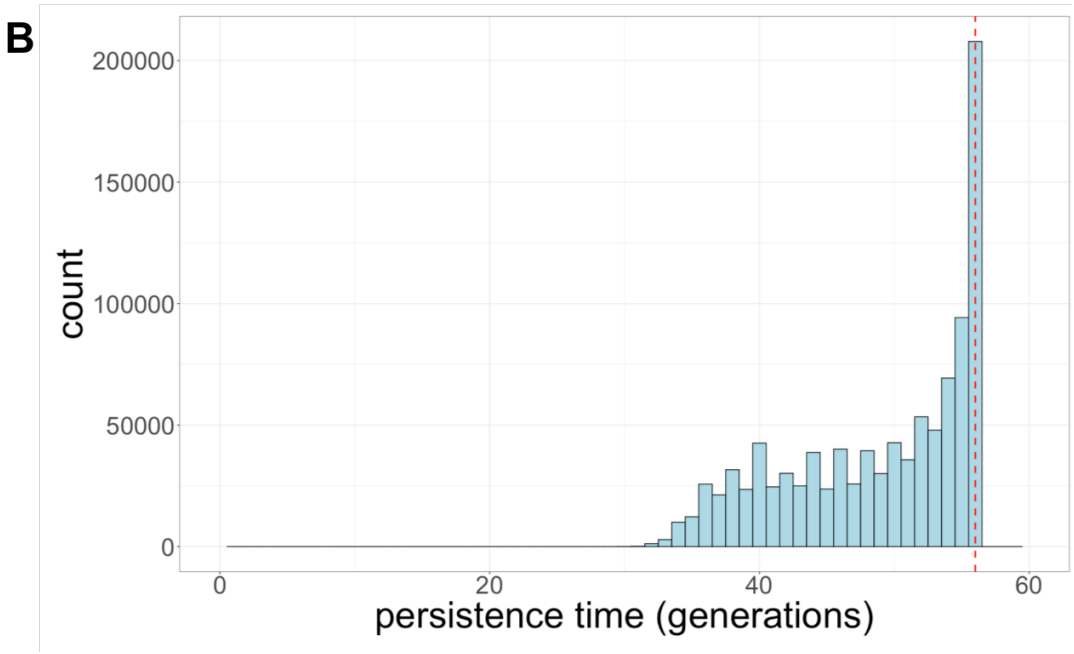
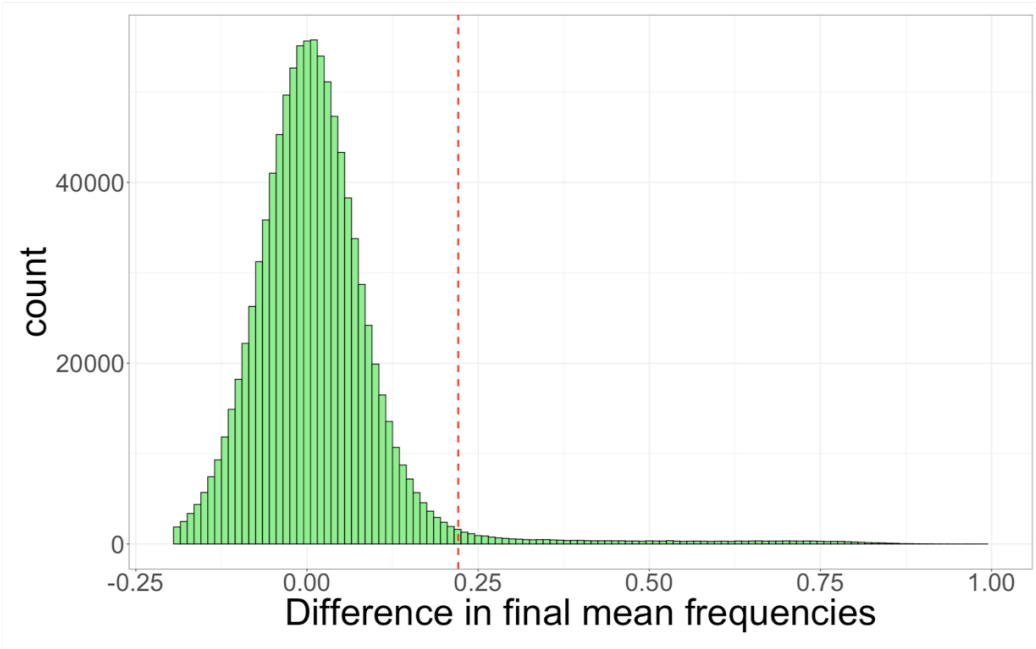


Figure C1. Selection of summary statistic values versus selection coefficient, *s*, for **A)** *positive directional selection*, and **B)** *negative directional selection* simulation models produced for SNP 1521. A random subset of 1000 simulation runs was used to create each set of plots. The true metric value for SNP 1521 is shown by the horizontal red line. In many of the comparisons the simulated summary statistic value flattens out at a value far from the SNP value once *s* is either >0.25 (**A**) or <-0.25 (**B**). This is due to rapid fixation of one or the other allele in most cages when selection is strong. Simulations with such strong values of *s* are unlikely to occur in reality and such runs are unhelpful in determining the mode of selection acting at the *fru* locus. From this observation it was decided to reduce the limits of *s* used in the simulations to a maximum value of 0.25.


A
60000
40000
20000
0
count
0
20
40
60
persistence time (generations)


B
200000
150000
100000
50000
0
count
0
20
40
60
persistence time (generations)

C


40000
20000
0
count
-0.25
0.00
0.25
0.50
0.75
1.00
Difference in final mean frequencies

D

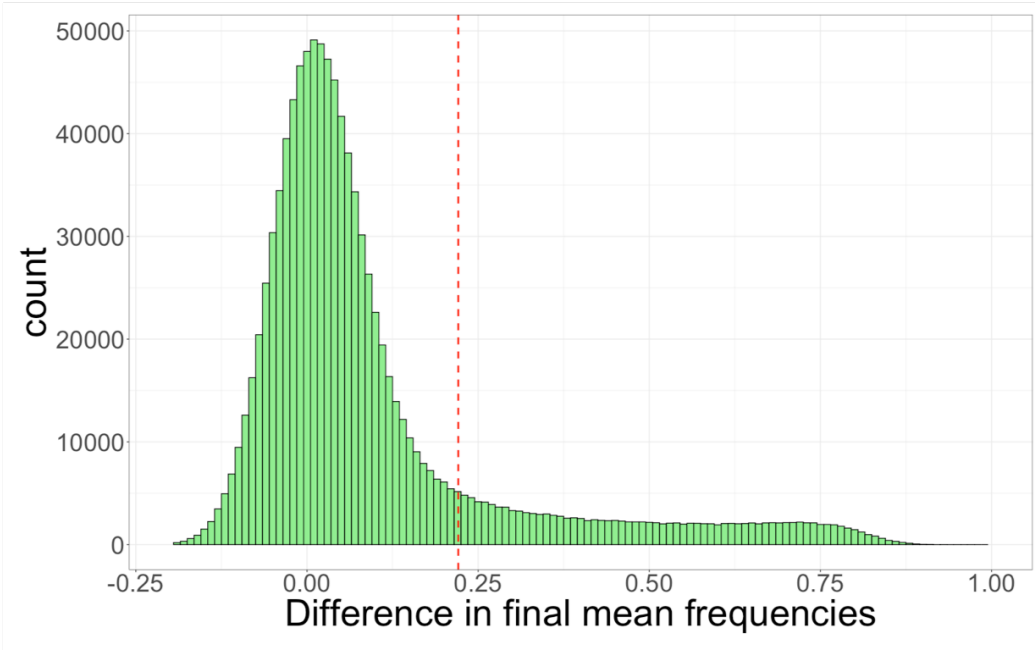
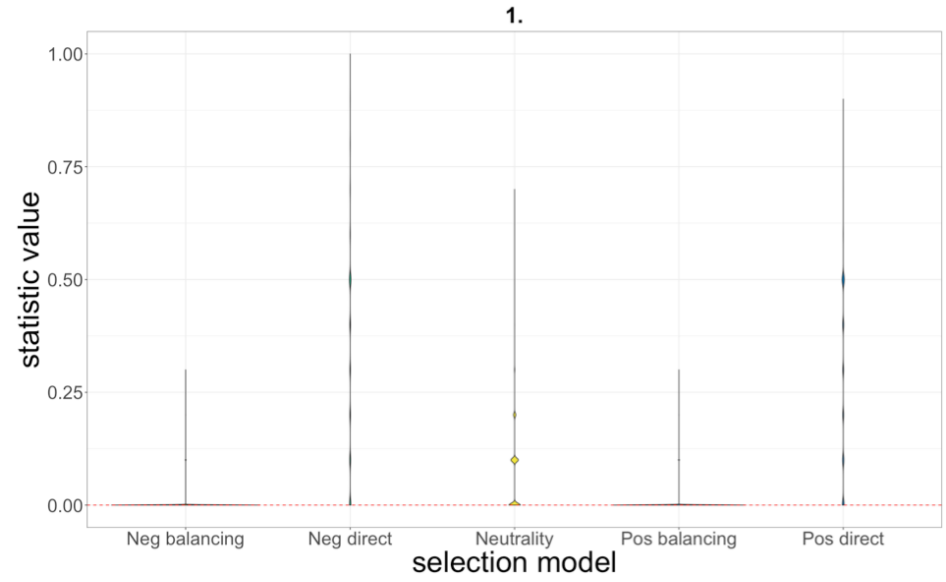
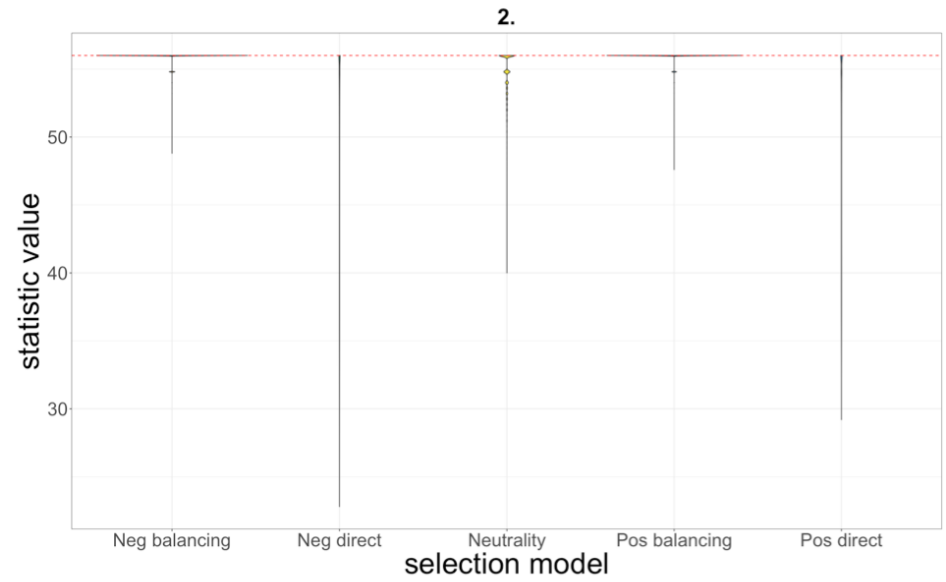
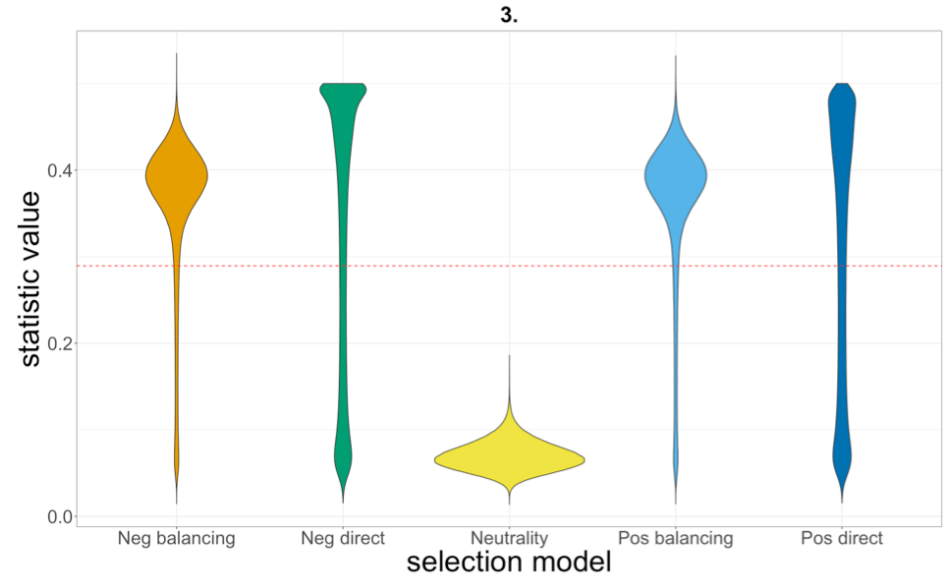
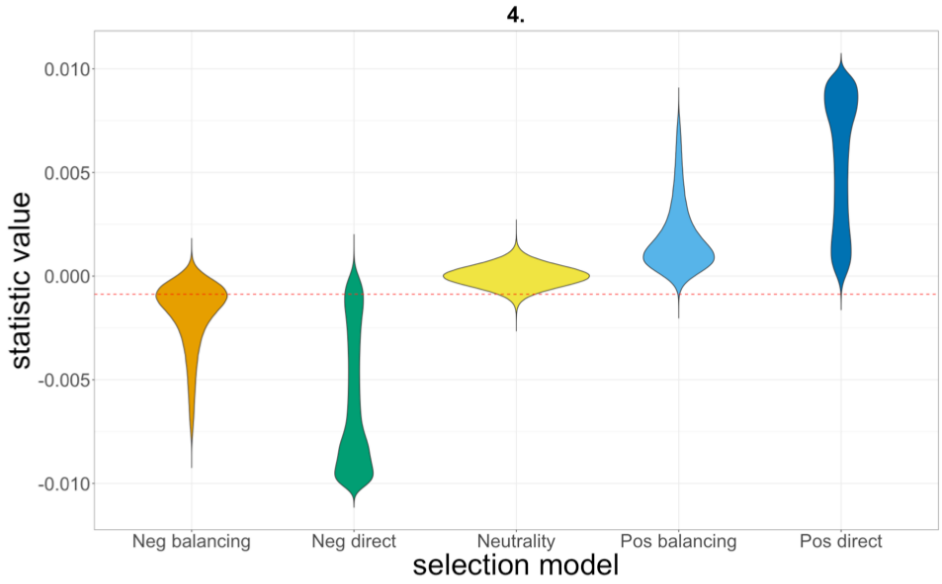

50000
40000
30000
20000
10000
0
count
-0.25
0.00
0.25
0.50
0.75
1.00
Difference in final mean frequencies

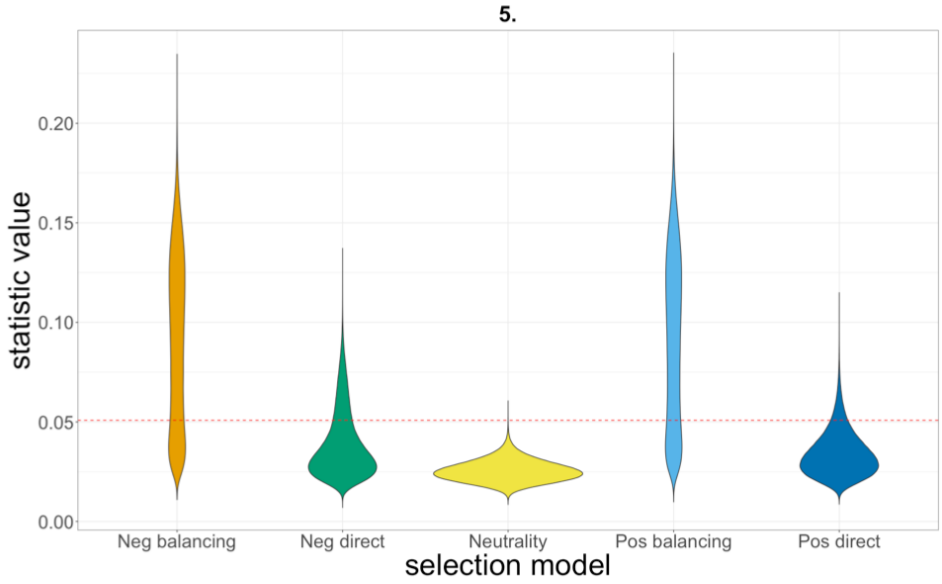
Figure C2. Example histograms produced to visualise simulation summary statistics for SNP 1521. **A** and **B** show a histogram of data collected for statistic 2, persistence time, for the *positive directional selection* model. The true SNP value is shown by the red dashed line. **A** presents data before the refinement of selection and parameters and **B** is after. Although even after refinement, this statistic does not show a high degree of overlap with the SNP value (Table C2) it does display a better fit overall than before. Histograms **C** and **D** show histograms for summary statistic 13, the difference in the final value means for the *positive balancing selection* model. The true SNP value is shown by the red dashed line. **C** presents data before the refinement of summary statistics, and **D** is after. In **C** there is little overlap with the data and the true value, but in **D** more of the data overlaps. This is observed in Table C2 where after refinement the degree of overlap between this statistic and the true value presents a good fit (i.e. >5% and <95% of values being larger). These histograms are examples. Equivalent plots were produced for all SNPs and selection models before and after the refinement of selection parameters to see how the changes affected the simulated data and to curate the summary statistics for the analysis.

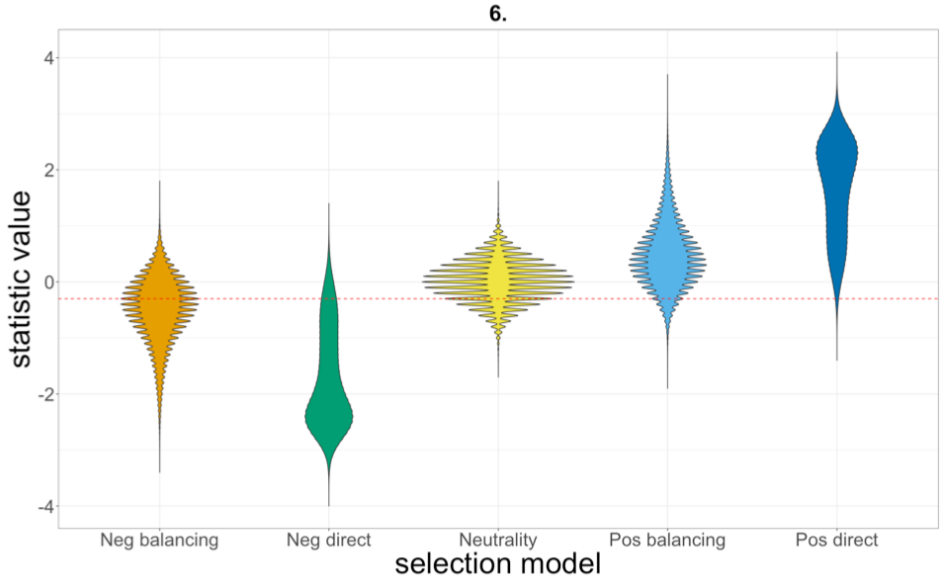

1.
1.00
0.75
0.50
0.25
0.00
statistic value
Neg balancing
Neg direct
Neutrality
Pos balancing
Pos direct
selection model

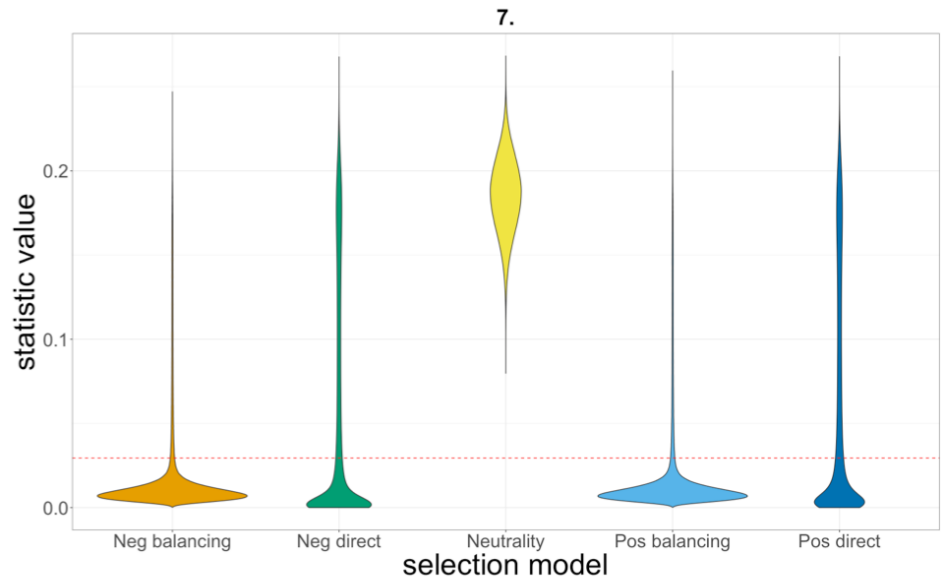

2.
50
40
30
statistic value
Neg balancing
Neg direct
Neutrality
Pos balancing
Pos direct
selection model

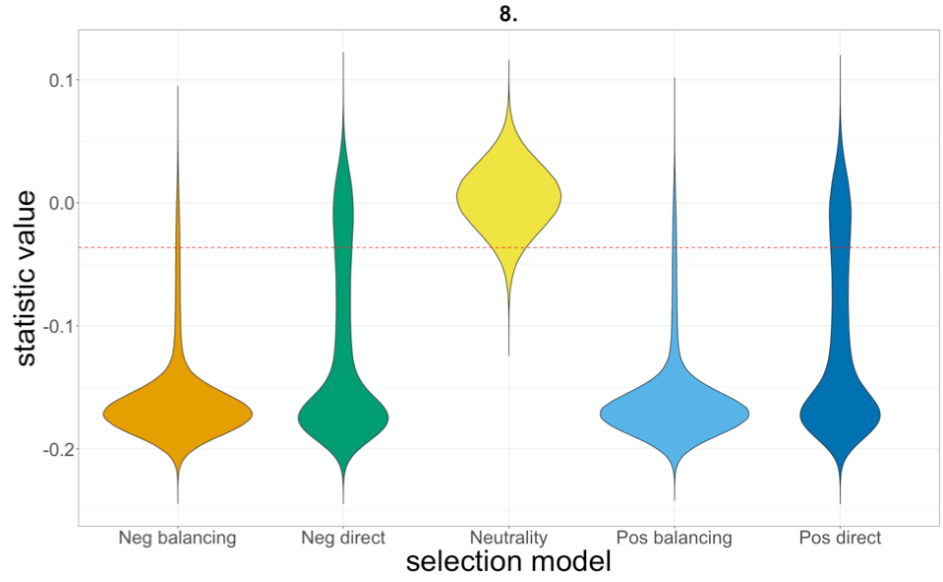

3.
0.4
0.2
0.0
statistic value
Neg balancing
Neg direct
Neutrality
Pos balancing
Pos direct
selection model

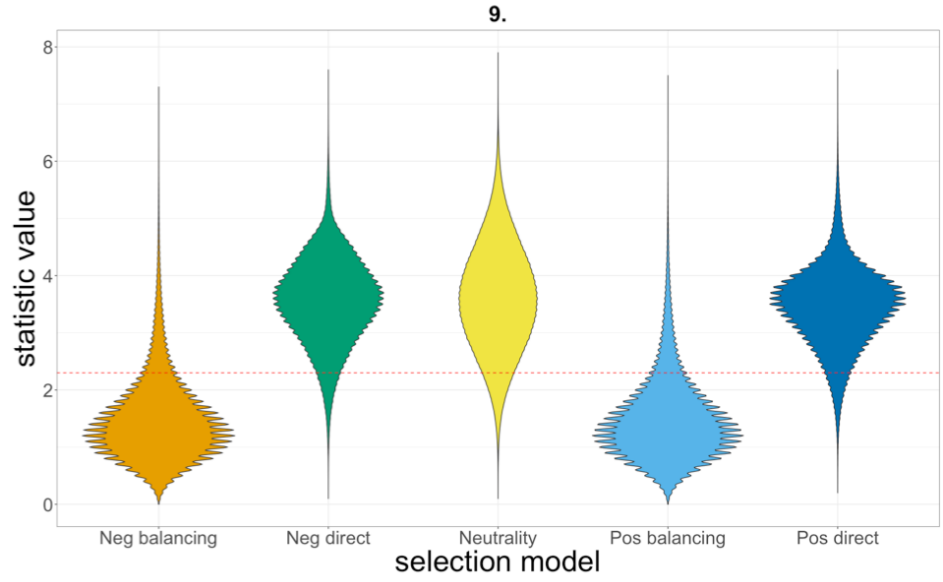

4.
0.010
0.005
0.000
-0.005
-0.010
statistic value
Neg balancing
Neg direct
Neutrality
Pos balancing
Pos direct
selection model


5.
0.20
0.15
0.10
0.05
0.00
statistic value
Neg balancing
Neg direct
Neutrality
Pos balancing
Pos direct
selection model

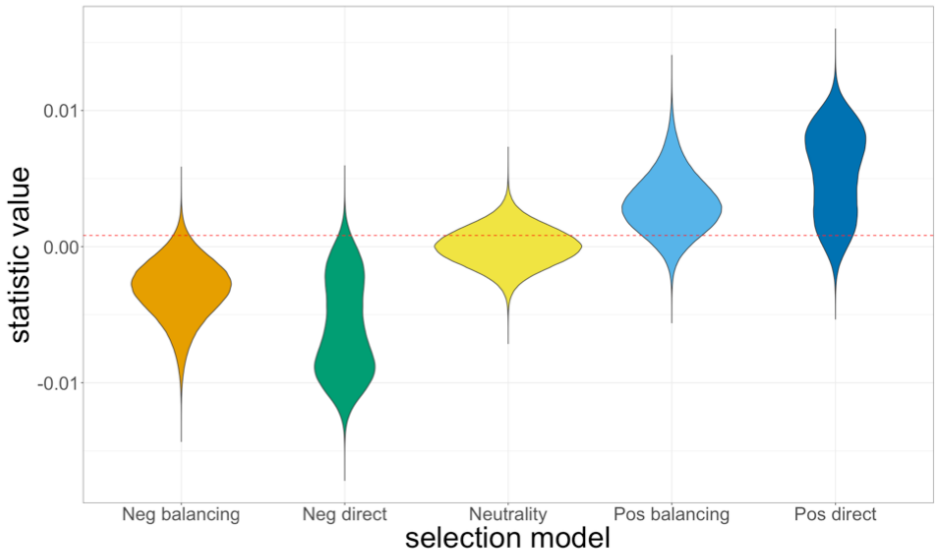

6.
4
2
0
-2
-4
statistic value
Neg balancing
Neg direct
Neutrality
Pos balancing
Pos direct
selection model


7.
statistic value
0.2
0.1
0.0
Neg balancing
Neg direct
Neutrality
Pos balancing
Pos direct
selection model

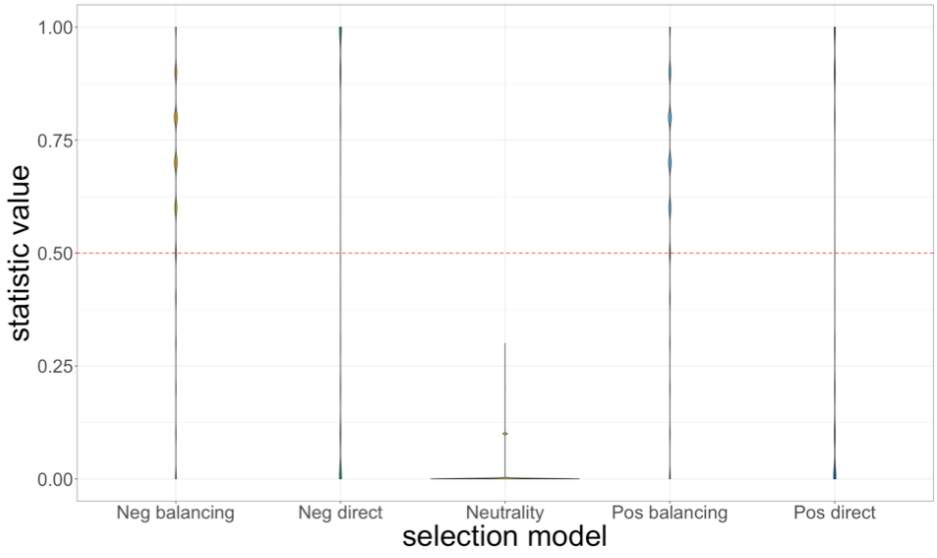

8.
statistic value
0.1
0.0
-0.1
-0.2
Neg balancing
Neg direct
Neutrality
Pos balancing
Pos direct
selection model


9.
statistic value
8
6
4
2
0
Neg balancing
Neg direct
Neutrality
Pos balancing
Pos direct
selection model

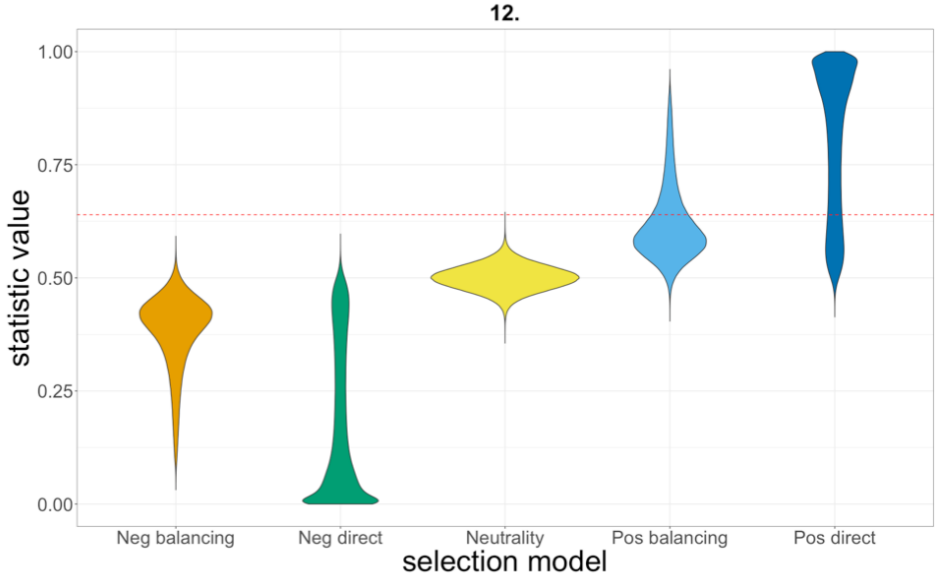
10.


0.01
0.00
-0.01
statistic value
Neg balancing
Neg direct
Neutrality
Pos balancing
Pos direct
selection model

11.


1.00
0.75
0.50
0.25
0.00
statistic value
Neg balancing
Neg direct
Neutrality
Pos balancing
Pos direct
selection model

12.


1.00
0.75
0.50
0.25
0.00
statistic value
Neg balancing
Neg direct
Neutrality
Pos balancing
Pos direct
selection model

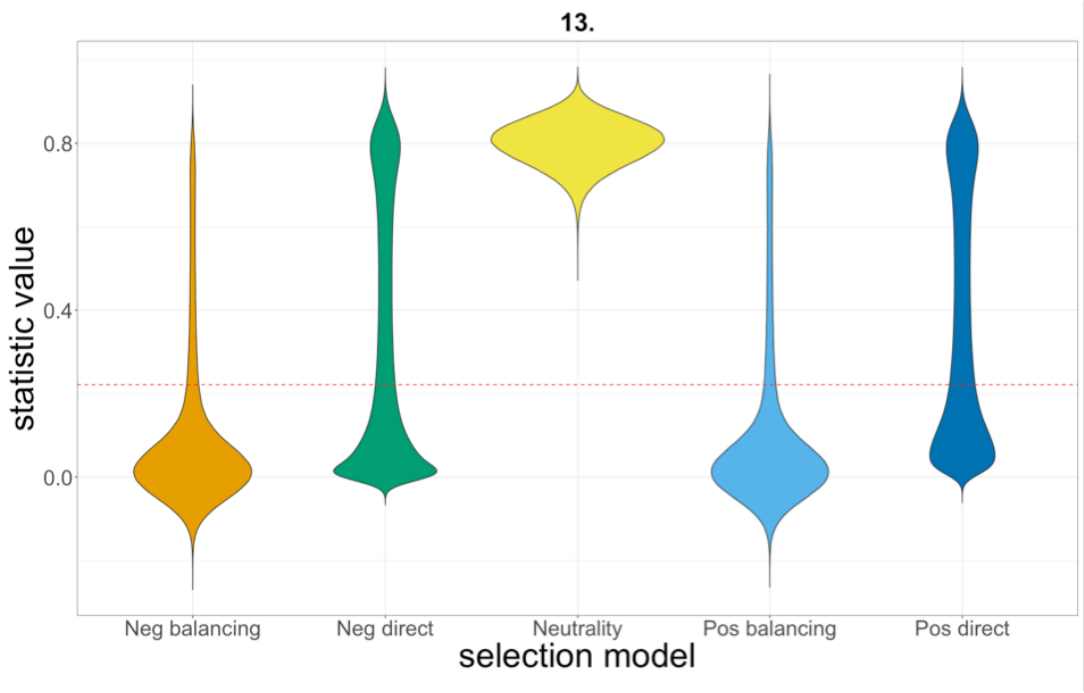


Figure C3. Violin plots of summary statistics per selection model. Each plot shows a different summary statistic denoted by the number above the plot. Numbers are those designated in Chapter 3. The summary statistic value for SNP 1521 is shown as dashed red horizontal line. Selection model names have been shortened for space so Neg = negative and Pos =positive. From these plots we can see the distribution of the statistic values vary between the different summary statistics. Even when two models display a similar distribution of values for one statistic they are different in another one. For example, in number 3 both *positive* and *negative balancing selection* have similar distributions, but in number 4 they are quite different. These differences are essential for ABC to tell the models apart and diagnose the mode of selection acting at the *fru* locus in Chapter 3.

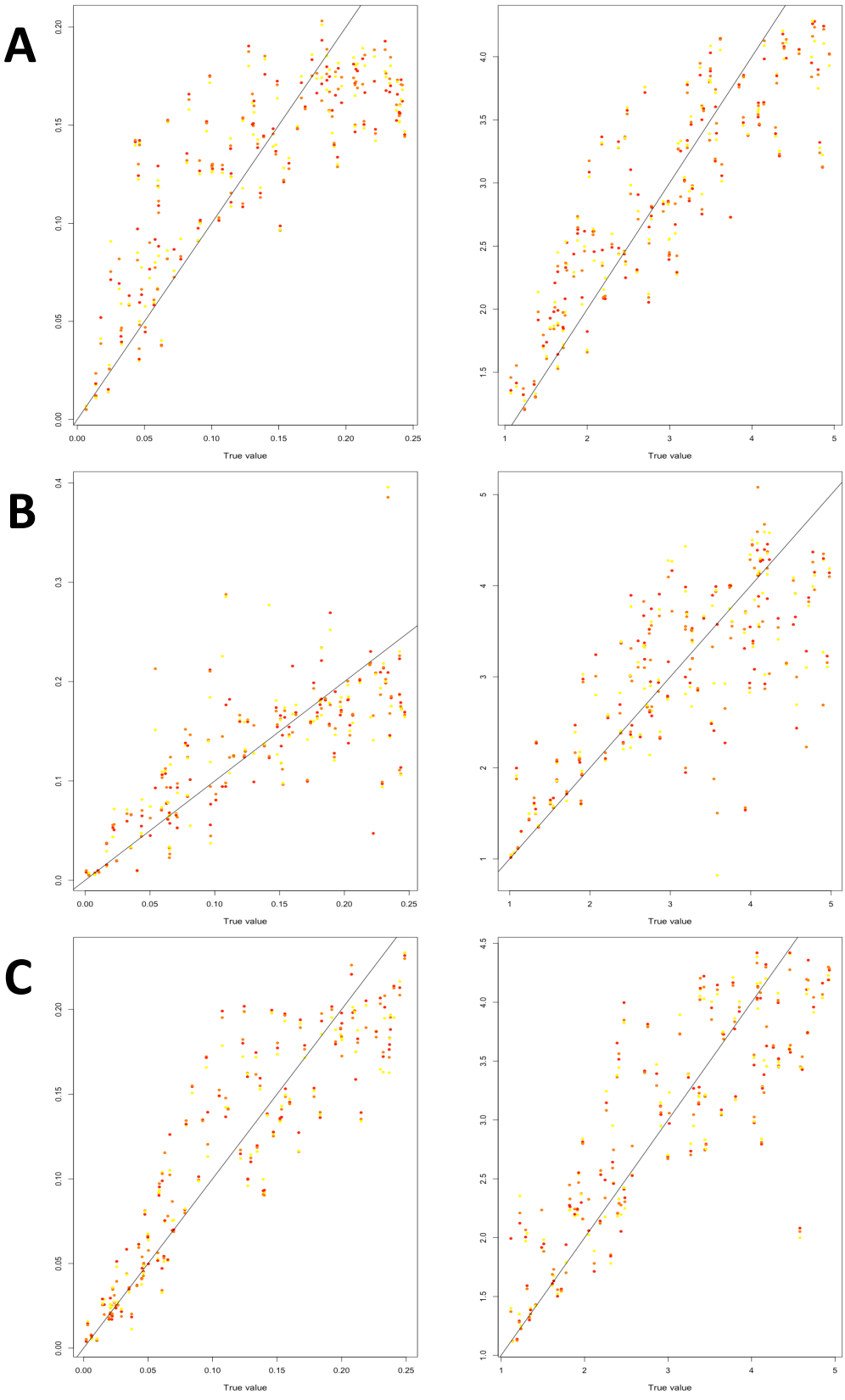

A
B
C
True value
True value
True value
True value
True value
True value

Figure C4. Parameter estimation cross-validation. The results from three different methods are shown: **A** = rejection; **B** = loc-linear; **C** = neural net Each simulation's true value of *s* and *h* is plotted against the value predicted based on selection metrics alone. Estimation for *h* is worse at higher values in all three methods. Estimation of *s* is best with the neural net method at lower values, but this breaks down at larger values. Overall error rates are shown in Table C3.

selection model class	Metric Number												
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
SNP.1181_neutral	0.833	0.859	0	0.545	0.009	0.17	1	0.805	0.912	0.077	<0.001	0	1
SNP.1208_neutral	0.745	0.755	<0.001	0.987	<0.001	0.719	0.999	0.924	0.973	0.192	0.003	0.003	0.999
SNP.1521_neutral	0.474	0.486	0	0.948	<0.001	0.763	1	0.929	0.91	0.285	0	<0.001	1
SNP.1181_positive_selection	0.971	0.977	0.897	0.996	0.878	0.969	0.108	0.044	0.993	0.934	0.869	0.907	0.122
SNP.1181_negative_selection	0.99	0.991	0.908	0.008	0.67	0.002	0.093	0.084	0.984	0.001	0.886	0	0.107
SNP.1208_positive_selection	0.965	0.967	0.926	0.999	0.706	0.996	0.099	0.054	0.998	0.959	0.89	0.959	0.096
SNP.1208_negative_selection	0.982	0.982	0.932	0.075	0.639	0.026	0.087	0.116	0.995	0.004	0.902	<0.001	0.086
SNP.1521_positive_selection	0.94	0.941	0.893	0.999	0.754	0.997	0.127	0.051	0.983	0.975	0.849	0.941	0.141
SNP.1521_negative_selection	0.947	0.948	0.905	0.046	0.729	0.026	0.111	0.108	0.984	0.005	0.87	0	0.122
SNP.1181_PosS_balancing_selection	0.011	0.126	0.979	0.568	0.959	0.206	0.014	0.003	0.332	0.421	0.973	0.057	0.018
SNP.1181_negS_balancing_selection	0.015	0.085	0.979	0.495	0.963	0.144	0.014	0.003	0.199	0.019	0.974	0	0.018
SNP.1208_PosS_balancing_selection	0.011	0.081	0.988	0.954	0.96	0.617	0.012	0.004	0.318	0.604	0.986	0.315	0.012
SNP.1208_negS_balancing_selection	0.015	0.085	0.988	0.839	0.957	0.575	0.012	0.004	0.313	0.062	0.986	<0.001	0.012
SNP.1521_PosS_balancing_selection	0.01	0.125	0.976	0.878	0.981	0.651	0.071	0.004	0.051	0.711	0.723	0.157	0.025
SNP.1521_negS_balancing_selection	0.014	0.126	0.976	0.766	0.98	0.612	0.071	0.004	0.047	0.105	0.723	<0.001	0.025

Table C1. Proportion of overlap values between simulation summary statistics and SNP summary statistics values. Summary statistics are referred to by the number given during the descriptions (Chapter 3). The selection model class and the SNP it is applied to appear in the leftmost column. Cells are colour coded by if the proportion of overlap value. Green = good overlap – model is consistent with data, orange = little overlap – model unlikely to give rise to the data but possible, red = no overlap – the model is inconsistent with the data. *Neutrality* models have between 4 and 6 summary statistics with <0.001% overlap making neutrality an unlikely model to have caused the real data. All other models have between 4 and 8 summary statistics with good levels of overlap, but these vary between selection model types and SNPs. Values are reported to 3 decimal places.

selection model class	Metric Number													Total with good overlap
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	
SNP.1521_neutral	0.474	0.486	0.000	0.948	0.000	0.763	1.000	0.929	0.910	0.285	0.000	0.000	1.000	5
SNP.1521_positive_selection	0.940	0.942	0.893	1.000	0.754	0.997	0.127	0.051	0.983	0.975	0.849	0.941	0.141	7
SNP.1521_negative_selection	0.947	0.948	0.904	0.045	0.728	0.026	0.112	0.108	0.984	0.006	0.870	0.000	0.123	6
SNP.1521_PosS_balancing_selection	0.010	0.125	0.976	0.878	0.981	0.651	0.071	0.004	0.051	0.711	0.726	0.157	0.025	8
SNP.1521_negS_balancing_selection	0.014	0.126	0.976	0.766	0.980	0.612	0.071	0.004	0.047	0.105	0.726	0.000	0.025	6
SNP.1521_pos_lower_selection	0.789	0.794	0.574	1.000	0.084	0.990	0.509	0.202	0.933	0.901	0.398	0.764	0.564	9
SNP.1521_neg_lower_selection	0.796	0.800	0.617	0.082	0.204	0.072	0.448	0.187	0.937	0.022	0.478	0.000	0.492	9
SNP.1521_bal_restricted_range_POS	0.010	0.039	0.857	0.998	0.775	0.911	0.128	0.024	0.111	0.862	0.743	0.298	0.151	11
SNP.1521_bal_restricted_range_NEG	0.007	0.039	0.903	0.443	0.849	0.438	0.087	0.016	0.095	0.025	0.800	0.000	0.102	10

Table C2. Proportion of overlap values for SNP 1521 including the five initial models, *neutrality*, *positive directional*, *negative directional*, *positive balancing*, and *negative balancing selection*, plus the values for the four models with the refined parameters ranges for *positive directional*, *negative directional*, *positive balancing*, and *negative balancing selection*. The final column shows the number of metrics that report a good proportion of overlap. The decision to refine the parameters is justified as every refined model has more metrics agreeing than its equivalent un-refined model. The refined parameter *positive balancing selection* model the best fitting based on these results alone with 11 out of 13 metrics agreeing. However, this is only one better than *negative balancing selection*, and 2 better than both *positive* and *negative directional selection* . Values are reported to 3 decimal places.

SNP	Tolerance (%)	Selection coefficient (*s*)	Dominance coefficient (*h*)
1181	0.5	0.173	0.289
	1	0.167	0.289
	5	0.192	0.313
1208	0.5	0.224	0.338
	1	0.225	0.347
	5	0.22	0.337
1521	0.5	0.124	0.221
	1	0.129	0.225
	5	0.124	0.221

Table C3. Cross-validation error in parameter estimation. Cross-validation performed with a sample size of 100 per model and tolerance limit. Values are reported to 3 decimal places.