

Modelling the formation of heterogeneous genomic landscapes



Hisham A.A. Ali
New College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Michaelmas 2023

Declaration

I declare that this thesis was composed by myself and that the work contained herein is my own except where explicitly stated in the text. The work has not been submitted for any degree or professional qualification except as specified.

Hisham A.A. Ali

Acknowledgments

My DPhil would not have been possible without the help of my friends and family. So, I would like to start this DPhil by thanking all those who have been alongside me.

Firstly, I would like to thank my primary supervisor Sonya Clegg. Thank you for your generosity with your advice and time through every step of my DPhil, from supporting my application to Oxford, to guiding me to my academic career after my DPhil. Secondly, I would like to thank Claudio Quilodr  n. Thank you for being a constant source of positive energy and encouragement which helped me along my DPhil, and for all your advice in coding and with the technical parts of the model. Thirdly, I would like to thank Tim Coulson, for always having an open door and providing invaluable advice. Thank you for guiding me through the challenges of the DPhil, and helping me to always find the solution. I could not have asked for better supervisors, whose kindness and encouragement guided me through my DPhil.

I also would like to thank all the members of the Clegg and Coulson lab, who have made my time at Oxford a joy. I would especially like to thank Andrew Wood and Tabitha Taberer for making sure our row was the best, and always putting a smile on my face. I would also like to thank my dear friends Ernest Wu, James Kempton and Alexandre Figueiredo for their unwavering support and kindness through my DPhil.

Last, and most importantly I would like to thank my family, without whom I would not be here. Thank you for listening to all my stories, and making my longest one possible. I would like to thank my parents. Without Ammi and Abbu as role models, who put me and my siblings first, I could never have hoped to come close to where I am today. I would like to thank Umaymah, Shayban and Hibban for always being there for me, from late night chats to ‘relaxing’ over a lightly competitive game of football. I would also like to thank Mum and Papa for their unwavering support and kindness through my DPhil. Lastly, I would like to thank my wife, Saba. When the road looks rough ahead, you are my anchor. Thank you for being ‘the friend in me’ for every day of my DPhil, through every up and down. Without your love and patience this journey would not have been possible.

Publications and Contributions

The following chapters are manuscripts that are either published or in preparation for submission. Each of the chapters are a collaborative effort with the following contribution. H.A.A.A., T.C., S.M.C., and C.S.Q. conceived the original idea for the study. H.A.A.A. coded and performed the simulations in R. H.A.A.A. wrote the first draft of the manuscript and all authors provided comments and edits on manuscript versions.

Chapter 2. Ali, H. A. A., Coulson, T., Clegg, S. M., & Quilodrán, C. S. (2023). The effect of divergent and parallel selection on the genomic landscape of divergence. *Molecular ecology*, 10.1111/mec.17225. Advance online publication. <https://doi.org/10.1111/mec.17225>

Chapter 3. Ali, H. A. A., Coulson, T., Clegg, S. M., & Quilodrán, C. S. The effect of Dobzhansky-Muller incompatibilities on the genomic landscape of divergence under parallel and divergent selection. *In preparation*.

Chapter 4. Ali, H. A. A., Coulson, T., Clegg, S. M., & Quilodrán, C. S. Effects of selection and gene flow barriers on the genomic landscape of introgression *In preparation*.

Abstract

The progression along the speciation continuum is driven by multiple evolutionary processes: selection, mutation, gene flow and drift. How these evolutionary forces, along with the genetic architecture e.g. recombination, affect the genomic landscape is a key question in evolutionary biology. The genomic landscape of divergence and introgression describe the heterogenous distribution of divergence and introgression along genomes, respectively. However, discerning which evolutionary processes shaped the evolution of a particular biological system is challenging. This is due to the fact that i) different strengths of the same evolutionary process can generate different progressions along the speciation continuum, and ii) through interaction between evolutionary processes, different evolutionary processes can produce that same pattern in the genomic landscape. Nevertheless, examining patterns in the genomic landscape can provide insights into the evolution of a particular biological system. In this thesis, I further develop an individual-based model to simulate the genomic landscape of divergence and introgression. In Chapter 2 and 3, I show the impact of additional layers of complexity on the genomic landscape of divergence. In Chapter 2, I explore the effect of i) different genotype-phenotype maps, by simulating genetic dominance and genotype-environment interaction; ii) frequency-dependent selection; and iii) varying the pattern of migration between populations. In Chapter 3, I model negative epistatic interactions in the form of Dobzhansky-Muller incompatibilities, and I introduce the simulation of multiple traits under selection. Through these developments, I emphasize

that the genomic landscape of divergence is highly influenced by additional layers of complexity, which must be considered when interpreting genomic patterns that emerge between diverging lineages. Additionally, I highlight that under the same evolutionary parameters, different genomic landscapes of divergence can form. In Chapter 4, I develop the individual-based model to examine the genomic landscape of introgression. Here I demonstrate the genomic landscape of introgression is highly variable, and even under the same evolutionary conditions very different genomic landscapes of introgression can form. Additionally, I highlight that concurrent interactions between evolutionary processes and the genomic architecture can change the effect an evolutionary process has on the genomic landscape. Our model provides a set of tools for better understanding the effects of different evolutionary processes concurrently and independently. Here, we show that the same evolutionary parameters can form different patterns on both the genomic landscape of divergence and introgression. Thus, caution must be taken in interpreting the genomic landscapes, as specific patterns in the genomic landscape are not necessarily indicative of a particular set of evolutionary conditions.

Table of Contents

Modelling the formation of heterogeneous genomic landscapes	i
Declaration.....	iii
Acknowledgments	iv
Publications and Contributions	vi
Abstract.....	vii
Table of Contents	ix
1. General Introduction	1
1.1 A brief introduction to evolutionary divergence	1
1.2 Methods of reconstructing evolutionary processes	3
1.3 The genomic landscapes of divergence and introgression	5
1.4 Modelling the genomic landscape.....	7
1.5 Aims of thesis	9
1.6 References	12
2. The effect of divergent and parallel selection on the genomic landscape of divergence	
.....	23
Abstract	23
2.1. Introduction	24
2.2. Methods	30
2.3. Results	40
2.4. Discussion	44
2.5. Conclusion.....	49
2.6. References	51

2.7. Figures and Tables.....	64
2.8 Supporting Information	76
3. The effect of Dobzhansky-Muller incompatibilities on the genomic landscape of divergence under parallel and divergent selection	84
Abstract	84
3.1. Introduction	85
3.2. Methods	88
3.3. Results	97
3.4. Discussion	99
3.5. Conclusion.....	103
3.6. References	104
3.7. Figures	110
3.8. Supporting information	116
4. Effects of selection and gene flow barriers on the genomic landscape of introgression	123
Abstract	123
4.1. Introduction	124
4.2. Methods	127
4.3. Results	134
4.4. Discussion	139
4.5. Conclusion.....	145
4.6. References	146
4.7. Figures	151
4.8 Supporting Information	157
5. Discussion	161

5.1 General discussion.....	161
5.2 An improved general model for simulating genomic landscapes	161
5.3 Individual based modelling	162
5.4 The variability of the genomic landscape.....	164
5.4.1 Genomic landscape of divergence.....	165
5.4.2 Genomic landscape of Introgression	168
5.5 Conclusion.....	169
5.5 References	171
A. Supplementary material: Chapter 2	175
Appendix 1	175
B. Supplementary material: Chapter 3	197
Appendix 1	197
C. Supplementary material: Chapter 4	228
Appendix 1	228
D. Supplementary material: Genomic landscapes of divergence among island bird populations: Evidence of parallel adaptation but at different loci?	249

1. General Introduction

1.1 A brief introduction to evolutionary divergence

The origin of species was referred to by Darwin as the “mystery of mysteries” (Darwin, 1859). The speciation process can be either a relatively abrupt event or a more gradual process (Abbott et al., 2016; Lamichhaney et al., 2018; Mavárez et al., 2006; Schwarz et al., 2005). The gradual process is often referred to as a speciation continuum, which depicts lineages diverging into predominantly reproductively isolated entities (Nosil, Harmon, et al., 2009). Theoretically, multiple evolutionary processes can drive progression along this continuum, which remains challenging to reconstruct for any particular case, especially after speciation. Studies of divergence below the species level can improve our understanding of the relative roles and interactions of evolutionary processes, which can promote divergence and may lead to species formation. One approach is to examine changes at the level of the genome for populations at different stages of divergence, and from this infer which evolutionary processes are in play (Ravinet et al., 2017). However, for this to be effective, we need to have a nuanced understanding of the types of genomic signatures that could be generated under different scenarios, a problem best suited to a modelling approach.

How the evolutionary processes: drift, selection, mutation, and gene flow each contribute to divergence along the speciation continuum is theoretically well understood (Yeaman et al., 2016). In essence, mutation can generate standing genetic variation, and these evolutionary processes can influence allele frequencies between populations, driving and hindering reproductive isolation, and in turn divergence. Divergence can be

driven by the neutral evolutionary process of genetic drift. Genetic drift is typically gradual and may act in tandem or against selection (Coyne & Orr, 2004a). However, the effects of drift may be significantly accelerated during a founder effect event or population bottleneck, during which profound changes in allele frequencies can be rapidly achieved (Coyne & Orr, 2004a). During a population bottleneck or founder event, the population size is very small, and so as explained by neutral theory, drift becomes a more powerful evolutionary force than natural selection, which can only overcome drift if it is very strong. Specifically, when the product of selection and effective population size ($N_e s$) is less than one, the mutation is effectively neutral (Kimura, 1979). Thus, drift can generate profound changes in deleterious alleles, even to the point of fixation. The evolutionary process of selection influences allele frequencies, by increasing the frequency of variants that confer a greater fitness in a given environment. Selection acts on genetic variation from either standing variation or novel mutations (Burke, 2012). When different variants are favoured between populations, it leads to divergence and potentially speciation (Rundle & Nosil, 2005). Gene flow is the movement of alleles between populations, which can hinder divergence by homogenising the populations (Nosil, 2008). However, gene flow between species via introgressive hybridisation is considered essential for some adaptive events (Lamichhaney et al., 2015; Mallet, 2007; Mavárez et al., 2006). Whilst these evolutionary processes have their own effects, in reality these processes may also act concordantly (Wolf & Ellegren, 2017). The mechanisms of drift and selection may act antagonistically or in tandem, favouring either different alleles or the same allele. Furthermore, the presence of gene flow homogenising genetic differentiation, may prevent divergence between populations. However, if the effect of drift and/or selection are stronger, then divergence will proceed. Thus, the presence of the same evolutionary processes at different relative strengths can produce

vastly different progressions along the speciation continuum. This is demonstrated by using sub-species and species complexes, e.g. *Heliconius* butterflies, the *Corvus* species complex, and silvereyes (*Zosterops lateralis*), as proxies for different stages along the speciation continuum (Martin et al., 2013; Sendell-Price et al., 2020; Vijay et al., 2016). In these comparisons between closely related species and populations, different relative strengths of microevolutionary processes have resulted in different evolutionary trajectories between the species and sub-species populations. For example, in *Heliconius* butterflies, decreasing the rate of gene flow corresponds with an increase in colour pattern and genic divergence (Supple et al., 2015).

1.2 Methods of reconstructing evolutionary processes

Phenotypic data can be used to provide an insight into the role of different evolutionary processes during divergence. For systems where sufficient information on population history is known, rate tests can be used to test the hypothesis: can the rate of phenotypic change be explained by drift alone (Lande, 1976). Several studies have used this approach and concluded that directional selection must be operating in addition to drift, to explain the rate and magnitude of divergence observed (Clegg, Degnan, Moritz, et al., 2002; Koskinen et al., 2002; Roberge et al., 2007). However, a limitation of this approach is that it requires large amounts of information, including divergence time, effective populations size and details of mutation rate, which are often not available in natural systems. Another approach is to examine the occurrence of parallel evolution, whereby populations evolve along parallel phenotypic trajectories. Typically, this suggests that natural selection has a role in divergence, since neutral processes are unlikely to explain recurring trends. Numerous examples in the literature have used parallel speciation as

evidence for the action of natural selection (Allender et al., 2003; Cai et al., 2019; Johannesson, 2001; Ostevik et al., 2012).

Genomic data can also be used to provide an insight into the role of different evolutionary processes during divergence. With the advent of next generation sequencing and computational advances, approaches have shifted from looking at a few genes to looking at genome wide patterns (Nosil & Feder, 2012). Several genome wide approaches test the hypothesis: is selection present. Firstly, the approach developed by McDonald and Kreitman (1991) compares polymorphisms with synonymous and non-synonymous amino acid sites, where selection is indicated by a difference in the nonsynonymous to synonymous ratios between species and within a species. Secondly, another approach for detecting selection, is by scanning the genome for evidence of a selective sweep, whereby an advantageous SNP tends to fixation, reducing genetic diversity at its position and in surrounding regions due to genetic hitch-hiking (Avalos et al., 2017; Eyre-Walker, 2006; Schlötterer, 2003). Thirdly, neutral equilibrium model-based tests predict genomic patterns that are anticipated under neutral evolution, which can then be compared to empirical data (Akey et al., 2004; Kern & Hahn, 2018). However, these approaches have limitations that require attention when interpreting results. For example, these models are susceptible to demographic effects e.g. demographic effects can mimic the effect of genomic hitch-hiking (Schlötterer, 2003; Eyre-Walker, 2006). Typically, these methods have prerequisites that empirical datasets cannot always satisfy, for example the approach developed by McDonald and Kreitman (1991) requires large datasets, and scanning the genome for selective sweeps is only effective when recent strong positive selection has occurred. Another challenge associated with genome wide approaches, rather than candidate loci-based approaches, is that due to the effects of hitchhiking and linkage they are only able to identify the presence and approximate location of a selected loci.

Contemporary genomic techniques have also been combined with phenotypic data, to provide insights into the mechanisms of divergence. One approach, is to compare the rate of divergence at neutral genetic markers with the phenotype (Barrowclough, 1983). This approach looks at the concordance between neutral loci and morphological trait changes, to test whether the phenotypic divergence can be explained by neutral mechanisms, and has been utilised by several studies (Clegg, Degnan, Moritz, et al., 2002; Leinonen et al., 2008; Whitlock, 2008). In the case of a positive correlation, this indicates neutral evolution, while negative correlation suggests the presence of selection. However, this method has been criticised, since a combination of sampling error and relatively small shifts between neutral genetic markers and the phenotype, can erroneously imply the presence of selection (Pujol et al., 2008). Despite numerous well worked examples and available methodologies, reconstructing how these evolutionary processes have acted and interacted to produce particular divergent forms is generally challenging.

1.3 The genomic landscapes of divergence and introgression

One area of promise for understanding the evolutionary history of species (i.e. the evolutionary processes that shape the evolution of a species), is by examining the genomic landscape of divergence: the heterogeneous profile of differentiation along the genome of two diverging lineages (Gavrilets, 2014; Michel et al., 2010; Wolf & Ellegren, 2017). During divergence, the genome usually does not diverge uniformly, generating regions of high and low divergence (relative to background levels of divergence), termed genomic islands and valleys, respectively (Ravinet et al., 2017; Roesti et al., 2014a). Genomic signatures of different evolutionary processes may be observed in the topography of the genomic landscape of divergence. Therefore, by examining the genomic landscape of

divergence, it may be possible to infer the roles that different evolutionary processes played in divergence (Ravinet et al., 2017). However, a major challenge is that different processes and combinations of processes can produce similar genomic signatures (Quilodr  n, Ruegg, et al., 2020a). For example, fluctuating selection with quasi-neutrality generates a nearly identical site-frequency spectrum to neutrality i.e. genetic drift (Tier, 1982; Ye et al., 2023). Therefore, a more nuanced understanding of the relationship between different evolutionary processes and their associated genomic signatures is integral for using the genomic landscape as a diagnostic tool for understanding genomic divergence.

Another tool in understanding the evolutionary history of a species is by examining the genomic landscape of introgression. Many species hybridise with close relatives, with estimates of at least 10% of mammals and 25% of plants (Mallet, 2005). Hybridisation is a key factor in the progression along the speciation continuum, with highly variable effects ranging from species fusion to differentiation (Seehausen, 2004). Introgression describes the transfer of DNA between genetically distinct lineages (i.e. subspecies and species), as a consequence of fertile hybridisation (Quilodr  n, Montoya-Burgos, et al., 2020). The genomic landscape of introgression: the topography of levels of introgression, is also typically heterogeneous, and characterised by regions of elevated and reduced levels of introgression, islands and valleys of introgression respectively. Like the genomic landscape of divergence, different evolutionary processes shape the landscape of introgression in different ways (Martin & Jiggins, 2017). The challenges of utilising the genomic landscape of introgression to understanding the evolutionary history of a species are similar to that of the genomic landscape of divergence: that while different evolutionary processes have different effects on the landscape of introgression, their effects can change through the concurrent interaction with other evolutionary processes.

Thus, interpreting the landscape of introgression also remains challenging. Previous approaches have had success in inferring the evolutionary history of a species. For example, Juric et al. (2016) were able to infer the strength of selection against Neanderthal introgression, through modelling the genomic data, while Aeschbacher et al. (2017) was able to infer migration and the strength against migration in the *Mimulus* monkeyflower (S. H. Martin & Jiggins, 2017). However, the landscape of introgression and divergence together can provide a critical insight into the evolution of species.

1.4 Modelling the genomic landscape

Many models have been developed that are limited to simulating either a single bi-allelic locus or a small number of loci (Charlesworth et al., 1997; Feder et al., 2012; Feder & Nosil, 2009, 2010; Sedghifar et al., 2016). Furthermore, often a single parameter is used for selection, providing a static view of divergence (Feder et al., 2012; Feder & Nosil, 2009, 2010). While these approaches have achieved success in modelling evolutionary divergence, they are often limited by a uniform distribution of loci with a constant recombination rate (Flaxman et al., 2013). Models such as MetaPopGen have also implemented gene flow, but at the caveat of being restricted to one locus (Andrello & Manel, 2015). Two major classes of models that have been used, and discussed below in more detail, are the backwards in time coalescence-based models (CBMs) and forwards in time individual based models (IBMs).

Coalescence based models have also been used to look at evolutionary divergence. CBMs are backwards in time models, that derive genealogies of the simulated loci, using coalescence theory, as a product of the model parameters (Hoban et al., 2012). This approach has had success when studying neutral evolution (Lohse, 2017). CBMs

also have the advantage of being able to incorporate large amounts of data, potentially simulating entire chromosomes of a large population in a computational efficient framework (Haller et al., 2019; Kelleher et al., 2016). CBMs have also been developed to include gene flow and hard selection, which influences population size as a function of the mean population fitness (Haller & Messer, 2019). CBMs are generally less informative for populations that are composed of few individuals experiencing selection, or which have non-panmictic reproduction in comparison to the next class of model, individual based models (IBMs) (Currat et al., 2015). Other limitations of CBMs, such as modelling gene flow, exponential growth and gene-species tree discordance, exist but have been mitigated (Eckert & Carstens, 2008; Edwards et al., 2016; Liu et al., 2009; Stadler et al., 2015). Additionally, the efficiency of CBMs to correctly identify the species tree despite incomplete sorting has been evaluated (Eckert & Carstens, 2008; Edwards et al., 2016).

Individual based models (IBMs), or agent-based models, are a flexible modelling framework that can be applied to studies of evolutionary divergence. IBMs provide insights at the population level from the collective behaviours of individuals (DeAngelis & Grimm, 2014; Grimm et al., 2006). They are forward-in-time models, in contrast to backwards-in-time coalescence simulations, facilitating the study of genomic composition change, such as the ones induced by non-panmictic reproduction or selection (Hoban et al., 2012). The common characteristics of IBMs are that the simulated individuals are defined by state variables e.g. physiological traits or genetic attributes, which can change over time, and the fixed population behaviours e.g. average offspring number and migration rates. IBMs have been successfully developed to simulate whole populations, and a large amount of genetic information; this is owed, in part, to the exponential increase in computational power (Currat et al., 2015; Peng & Kimmel, 2005).

Additionally, some models have utilised a mixed forward-backward approach incorporating individual and coalescence-based aspects into the model, utilising the advantage of the speed of CBMs and the ability of IBMs to simulate selection (Currat et al., 2019; Haller & Messer, 2019). Whilst dealing with a different question, Smadbeck & Stumpf (2016) evaluated the relative performances of CBMs and IBMs for cell tracking in lineage in tissue growth.

The *glads* (*Genomic Landscapes of Divergence Simulations*) IBM developed by Quilodran et al. (2019), implements several major innovations for modelling genomic landscape. First, the model includes a flexible fitness function, linking evolution, in the form of selection, with ecology, in the form of population demographics. Furthermore, in contrast to the hard selection found in CBMs, population size is dependent on individual fecundity, a function of the individual's fitness, rather than the average fitness of the population. Second, the IBM facilitates various scenarios of gene flow and selection, on single or multiple interacting genes. Third, the simulation can incorporate any number of loci and polymorphisms in a large population.

1.5 Aims of thesis

With the developments in the field of modelling the evolution of the genome of species, restrictions remain in simulating selection and different genotype-phenotype maps. Here, I introduce additional layers of complexity to the *glads* model to incorporate different types of selection, genotype-phenotype maps and more freedom for the demographic migration parameter. These changes are incorporated with the aim of achieving a more complete understanding of the processes and scenarios that result in particular features of the genomic landscape of divergence. I also extend the model, with the aim of achieving

a more nuanced understanding of the genomic landscape of introgression. By utilising these modifications together, I aim to produce a set of tools that helps us understand the evolution of a particular biological system.

The Oceanic silvereye passerine (*Zosterops lateralis*) offers an excellent opportunity to study the speciation continuum. This island-colonising species complex has repeatedly colonised numerous islands in the region of its ancestral population on the Australian mainland (Mees, 1969). With divergence time frames ranging from <150 years to 100,000 years, this range of divergence time frames can be used as a proxy of the speciation continuum, illustrating incipient to late-stage divergence (Clegg, Degnan, Kikkawa, et al., 2002). Additionally, some silvereye populations have evolved with gene flow, while others have evolved without gene flow (Mees, 1969). The silvereyes have also shown a repeated increase in body size on different islands, relative to the ancestral mainland population (Clegg, Degnan, Moritz, et al., 2002; Clegg et al., 2008; Mees, 1969). However, bill shape has been found to evolve differentially across the island populations (Clegg & Phillimore, 2010), suggesting these two traits are evolving under parallel and divergent selection respectively. This combination of parallel and divergent selection, presence and absence of gene flow, and different points along the speciation continuum, makes the silvereye an excellent empirical system for modelling these different microevolutionary processes. Sendell-Price et al. (2020) used an individual based modelling approach to model the accumulation of genome wide divergence in the silvereyes, and observed under allopatry and low levels of gene flow a faster accumulation of genome wide divergence than observed empirically in the silvereyes. This approach highlights how these approaches can be utilised to understand the evolutionary history of a species, and how limitations and parametrisation based on empirical bench marking can be identified through comparison with empirical data.

In Chapter 2, I extend the *glads* model to look at the effects of additional layers of complexity on the genomic landscape of divergence. Firstly, I introduce both positive and negative frequency-dependent selection to the model. Secondly, I introduce ways to modify the genotype-phenotype map beyond a simple additive genotype-phenotype map, specifically examining the effects of genetic dominance and genotype-environment interaction on the genomic landscape. Thirdly, I modify the migration parameter to simulate variable patterns of migration rather than a fix rate over generations. In this chapter, the effects of each these modifications on the genomic landscape of divergence are examined concurrently with divergent and parallel selection, neutral evolution, and with or without linkage.

In Chapter 3, I extend the *glads* model to examine the effects of Dobzhansky-Muller incompatibilities mutations on the genomic landscape of divergence. In this chapter, I also examine the effects of concurrently simulating multiple traits under divergent and parallel selection on the genomic landscape. Lastly, I examine the effects of these modifications under different migration rates.

In Chapter 4, I extend the *glads* model to examine the genomic landscape of introgression. Firstly, I examine the effects of a reduced fitness of hybrids on the genomic landscape of introgression. Secondly, I introduce gene of large effects into the model and examine the effect of their density on the landscape of introgression. Thirdly, I examine the concurrent effect of these additions with different evolutionary processes, including a non-recombining sex chromosome.

The *glads* model was chosen to implement these additions for several reasons. Firstly, the *glads* model was developed to be easily modifiable, and so it was possible to implement not only the genotype-phenotype map, but also the modified migration, mutation and selection parameters (Quilodr  n, Ruegg, et al., 2020a). Secondly, the

programme was developed in the statistical programming language R (R Development Core Team, 2017). I was able to utilise my familiarity with the R program to implement the changes into *glads* and into the c++ functions that the *glads* package also used. Thirdly, while other approaches have been developed that allow the simulation of the loci under different evolutionary processes e.g. Slim, Splathe, and those discussed above in 1.4, the *glads* model unlike these models was developed to specifically to focus on the genomic landscape (Currat et al., 2019; Haller & Messer, 2019). At the start of this DPhil the authors of the *glads* model were looking to further develop the *glads* model. Many of the innovations that are available today for some of the other modelling approaches were developed after the start of this DPhil, while the research presented in this DPhil was being developed.

1.6 References

- Abbott, R. J., Barton, N. H., & Good, J. M. (2016). Genomics of hybridization and its evolutionary consequences. *Molecular Ecology*, 25(11), 2325–2332. <https://doi.org/10.1111/mec.13685>
- Aeschbacher, S., Selby, J. P., Willis, J. H., & Coop, G. (2017). Population-genomic inference of the strength and timing of selection against gene flow. *Proceedings of the National Academy of Sciences of the United States of America*, 114(27), 7061–7066. <https://doi.org/10.1073/pnas.1616755114>
- Akey, J. M., Eberle, M. A., Rieder, M. J., Carlson, C. S., Shriver, M. D., Nickerson, D. A., & Kruglyak, L. (2004). Population history and natural selection shape patterns of genetic variation in 132 genes. *PLoS Biology*, 2(10). <https://doi.org/10.1371/journal.pbio.0020286>

- Allender, C. J., Seehausen, O., Knight, M. E., Turner, G. F., & Maclean, N. (2003). Divergent selection during speciation of Lake Malawi cichlid fishes inferred from parallel radiations in nuptial coloration. *Proceedings of the National Academy of Sciences of the United States of America*, 100(SUPPL. 2), 14074–14079. <https://doi.org/10.1073/pnas.2332665100>
- Andrello, M., & Manel, S. (2015). MetaPopGen: An R package to simulate population genetics in large size metapopulations. *Molecular Ecology Resources*, 15(5), 1153–1162. <https://doi.org/10.1111/1755-0998.12371>
- Avalos, A., Pan, H., Li, C., Acevedo-Gonzalez, J. P., Rendon, G., Fields, C. J., Brown, P. J., Giray, T., Robinson, G. E., Hudson, M. E., & Zhang, G. (2017). A soft selective sweep during rapid evolution of gentle behaviour in an Africanized honeybee. *Nature Communications*, 8(1). <https://doi.org/10.1038/s41467-017-01800-0>
- Barrowclough, G. F. (1983). Biochemical studies of microevolutionary processes. *Perspective in Ornithology*, 223–261.
- Burke, M. K. (2012). How does adaptation sweep through the genome? Insights from long-term selection experiments. *Proceedings of the Royal Society B: Biological Sciences*, 279(1749), 5029–5038. <https://doi.org/10.1098/rspb.2012.0799>
- Cai, Z., Zhou, L., Ren, N. N., Xu, X., Liu, R., Huang, L., Zheng, X. M., Meng, Q. L., Du, Y. S., Wang, M. X., Geng, M. F., Chen, W. L., Jing, C. Y., Zou, X. H., Guo, J., Chen, C. Bin, Zeng, H. Z., Liang, Y. T., Wei, X. H., ... Ge, S. (2019). Parallel Speciation of Wild Rice Associated with Habitat Shifts. *Molecular Biology and Evolution*, 36(5), 875–889. <https://doi.org/10.1093/molbev/msz029>
- 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.
<https://doi.org/10.1017/S0016672397002954>
- Clegg, S. M., Degnan, S. M., Kikkawa, J., Moritz, C., Estoup, A., & Owens, I. P. F. (2002). Genetic consequences of sequential founder events by an island-colonizing bird. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12), 8127–8132. <https://doi.org/10.1073/pnas.102583399>
- Clegg, S. M., Degnan, S. M., Moritz, C., Estoup, A., Kikkawa, J., & Owens, I. P. F. (2002). Microevolution in island forms: The roles of drift and directional selection in morphological divergence of a passerine bird. *Evolution*, 56(10), 2090–2099.
<https://doi.org/10.1111/j.0014-3820.2002.tb00134.x>
- Clegg, S. M., Frentiu, F. D., Kikkawa, J., Tavecchia, G., & Owens, I. P. F. (2008). 4000 Years of phenotypic change in an island bird: Heterogeneity of selection over three microevolutionary timescales. *Evolution*, 62(9), 2393–2410.
<https://doi.org/10.1111/j.1558-5646.2008.00437.x>
- Clegg, S. M., & Phillimore, A. B. (2010). The influence of gene flow and drift on genetic and phenotypic divergence in two species of *Zosterops* in Vanuatu. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1543), 1077–1092.
<https://doi.org/10.1098/rstb.2009.0281>
- Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sinauer Associates, Inc. Publishers.
- Curat, M., Arenas, M., Quilodrán, C. S., Excoffier, L., & Ray, N. (2019). SPLATCHE3: Simulation of serial genetic data under spatially explicit evolutionary scenarios including long-distance dispersal. *Bioinformatics*, 35(21), 4480–4483.
<https://doi.org/10.1093/bioinformatics/btz311>

- Curat, M., Gerbault, P., Di, D., Nunes, J. M., & Sanchez-Mazas, A. (2015). Forward-in-time, spatially explicit modeling software to simulate genetic lineages under selection. *Evolutionary Bioinformatics*, 11, 27–39. <https://doi.org/10.4137/EBO.S33488>
- Darwin, C. (1859). On the origin of species by means of natural selection. In *Murray: London, UK*.
- DeAngelis, D. L., & Grimm, V. (2014). Individual-based models in ecology after four decades. *F1000Prime Reports*, 6(June). <https://doi.org/10.12703/P6-39>
- Eckert, A. J., & Carstens, B. C. (2008). Does gene flow destroy phylogenetic signal? The performance of three methods for estimating species phylogenies in the presence of gene flow. *Molecular Phylogenetics and Evolution*, 49(3), 832–842. <https://doi.org/10.1016/j.ympev.2008.09.008>
- Edwards, S. V., Xi, Z., Janke, A., Faircloth, B. C., McCormack, J. E., Glenn, T. C., Zhong, B., Wu, S., Lemmon, E. M., Lemmon, A. R., Leaché, A. D., Liu, L., & Davis, C. C. (2016). Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. In *Molecular Phylogenetics and Evolution* (Vol. 94, pp. 447–462). Academic Press Inc. <https://doi.org/10.1016/j.ympev.2015.10.027>
- Eyre-Walker, A. (2006). The genomic rate of adaptive evolution. *Trends in Ecology and Evolution*, 21(10), 569–575. <https://doi.org/10.1016/j.tree.2006.06.015>
- Feder, J. L., Gejji, R., Yeaman, S., & Nosil, P. (2012a). Establishment of new mutations under divergence and genome hitchhiking. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1587), 461–474. <https://doi.org/10.1098/rstb.2011.0256>
- Feder, J. L., Gejji, R., Yeaman, S., & Nosil, P. (2012b). Establishment of new mutations under divergence and genome hitchhiking. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1587), 461–474. <https://doi.org/10.1098/rstb.2011.0256>

- Feder, J. L., & Nosil, P. (2009). Chromosomal inversions and species differences: When are genes affecting adaptive divergence and reproductive isolation expected to reside within inversions? *Evolution*, 63(12), 3061–3075. <https://doi.org/10.1111/j.1558-5646.2009.00786.x>
- Feder, J. L., & Nosil, P. (2010). The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution: International Journal of Organic Evolution*, 64(6), 1729–1747. <https://doi.org/10.1111/j.1558-5646.2009.00943.x>
- Flaxman, S. M., Feder, J. L., & Nosil, P. (2013). Genetic hitchhiking and the dynamic buildup of genomic divergence during speciation with gene flow. *Evolution*, 67(9), 2577–2591. <https://doi.org/10.1111/evo.12055>
- Gavrilets, S. (2014). Models of speciation: Where are we now? *Journal of Heredity*, 105(S1), 743–755. <https://doi.org/10.1093/jhered/esu045>
- Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S. K., Huse, G., Huth, A., Jepsen, J. U., Jørgensen, C., Mooij, W. M., Müller, B., Pe'er, G., Piou, C., Railsback, S. F., Robbins, A. M., ... DeAngelis, D. L. (2006). A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*, 198(1–2), 115–126. <https://doi.org/10.1016/j.ecolmodel.2006.04.023>
- Haller, B. C., Galloway, J., Kelleher, J., Messer, P. W., & Ralph, P. L. (2019). Tree-sequence recording in SLiM opens new horizons for forward-time simulation of whole genomes. *Molecular Ecology Resources*, 19(2), 552–566. <https://doi.org/10.1111/1755-0998.12968>
- Haller, B. C., & Messer, P. W. (2019). SLiM 3: Forward Genetic Simulations Beyond the Wright-Fisher Model. *Molecular Biology and Evolution*, 36(3), 632–637. <https://doi.org/10.1093/molbev/msy228>

- Hoban, S., Bertorelle, G., & Gaggiotti, O. E. (2012). Computer simulations: Tools for population and evolutionary genetics. *Nature Reviews Genetics*, 13(2), 110–122. <https://doi.org/10.1038/nrg3130>
- Johannesson, K. (2001). Parallel speciation: A key to sympatric divergence. *Trends in Ecology and Evolution*, 16(3), 148–153. [https://doi.org/10.1016/S0169-5347\(00\)02078-4](https://doi.org/10.1016/S0169-5347(00)02078-4)
- Juric, I., Aeschbacher, S., & Coop, G. (2016). The Strength of Selection against Neanderthal Introgression. *PLoS Genetics*, 12(11). <https://doi.org/10.1371/journal.pgen.1006340>
- Kelleher, J., Etheridge, A. M., & McVean, G. (2016). Efficient Coalescent Simulation and Genealogical Analysis for Large Sample Sizes. *PLoS Computational Biology*, 12(5), 1–22. <https://doi.org/10.1371/journal.pcbi.1004842>
- Kern, A. D., & Hahn, M. W. (2018). The neutral theory in light of natural selection. *Molecular Biology and Evolution*, 35(6), 1366–1371. <https://doi.org/10.1093/molbev/msy092>
- Kimura, M. (1979). *Model of effectively neutral mutations in which selective constraint is incorporated (molecular evolution/protein polymorphism/population genetics/neutral mutation theory)* (Vol. 76, Issue 7). <https://www.pnas.org>
- Koskinen, M. T., Haugen, T. O., & Primmer, C. R. (2002). Contemporary fisherian life-history evolution in small salmonid populations. *Nature*, 419(6909), 826–830. <https://doi.org/10.1038/nature01029>
- Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., Promerová, M., Rubin, C. J., Wang, C., Zamani, N., Grant, B. R., Grant, P. R., Webster, M. T., & Andersson, L. (2015). Evolution of Darwin’s finches and their beaks revealed by genome sequencing. *Nature*, 518(7539), 371–375. <https://doi.org/10.1038/nature14181>
- Lamichhaney, S., Han, F., Webster, M. T., Andersson, L., Grant, B. R., & Grant, P. R. (2018). Rapid hybrid speciation in Darwin’s finches. *Science*, 359(6372), 224–228.

- Lande, R. (1976). Natural Selection and Random Genetic Drift in Phenotypic Evolution. *Evolution*, 30(2), 314. <https://doi.org/10.2307/2407703>
- Leinonen, T., O'Hara, R. B., Cano, J. M., & Merilä, J. (2008). Comparative studies of quantitative trait and neutral marker divergence: A meta-analysis. *Journal of Evolutionary Biology*, 21(1), 1–17. <https://doi.org/10.1111/j.1420-9101.2007.01445.x>
- Liu, L., Yu, L., Pearl, D. K., & Edwards, S. V. (2009). Estimating species phylogenies using coalescence times among sequences. *Systematic Biology*, 58(5), 468–477. <https://doi.org/10.1093/sysbio/syp031>
- Lohse, K. (2017). Come on feel the noise – from metaphors to null models. *Journal of Evolutionary Biology*, 30(8), 1506–1508. <https://doi.org/10.1111/jeb.13109>
- Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology and Evolution*, 20(5), 229–237. <https://doi.org/10.1016/j.tree.2005.02.010>
- Mallet, J. (2007). Hybrid speciation. *Nature*, 446(7133), 279–283. <https://doi.org/10.1038/nature05706>
- Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., Blaxter, M., Manica, A., Mallet, J., & Jiggins, C. D. (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research*, 23(11), 1817–1828. <https://doi.org/10.1101/gr.159426.113>
- Martin, S. H., & Jiggins, C. D. (2017). Interpreting the genomic landscape of introgression. *Current Opinion in Genetics and Development*, 47, 69–74. <https://doi.org/10.1016/j.gde.2017.08.007>
- Mavárez, J., Salazar, C. A., Bermingham, E., Salcedo, C., Jiggins, C. D., & Linares, M. (2006). Speciation by hybridization in *Heliconius* butterflies. *Nature*, 441(7095), 868–871. <https://doi.org/10.1038/nature04738>

- McDonald, J. H., & Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*, *351*(6328), 652–654.
- Mees, G. F. (1969). A systematic review of the Indo-Australian Zosteropidae (Part III). *Zoologische Verhandelingen*, *102*(1), 1–390.
- Michel, A. P., Sim, S., Powell, T. H. Q., Taylor, M. S., Nosil, P., & Feder, J. L. (2010). Widespread genomic divergence during sympatric speciation. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(21), 9724–9729. <https://doi.org/10.1073/pnas.1000939107>
- Nosil, P. (2008). Speciation with gene flow could be common. *Molecular Ecology*, *17*(9), 2006–2008. <https://doi.org/10.1186/1742-9994-4-11>.Schmitt
- Nosil, P., & Feder, J. L. (2012). Genomic divergence during speciation: Causes and consequences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1587), 332–342. <https://doi.org/10.1098/rstb.2011.0263>
- Nosil, P., Harmon, L. J., & Seehausen, O. (2009). Ecological explanations for (incomplete) speciation. *Trends in Ecology and Evolution*, *24*(3), 145–156. <https://doi.org/10.1016/j.tree.2008.10.011>
- Ostevik, K. L., Moyers, B. T., Owens, G. L., & Rieseberg, L. H. (2012). Parallel ecological speciation in plants? *International Journal of Ecology*, *2012*. <https://doi.org/10.1155/2012/939862>
- Peng, B., & Kimmel, M. (2005). simuPOP: A forward-time population genetics simulation environment. *Bioinformatics*, *21*(18), 3686–3687. <https://doi.org/10.1093/bioinformatics/bti584>
- Pujol, B., Wilson, A. J., Ross, R. I. C., & Pannell, J. R. (2008). Are QST-FST comparisons for natural populations meaningful? *Molecular Ecology*, *17*(22), 4782–4785. <https://doi.org/10.1111/j.1365-294X.2008.03958.x>

- Quilodran, A. C. S., Ruegg, K., Sendell-, A. T., Clegg, S., Coulson, T., Anderson, E., & Quilodran, M. C. S. (2019). *Package 'glads.'* 1–13.
- Quilodrán, C. S., Montoya-Burgos, J. I., & Currat, M. (2020). Harmonizing hybridization dissonance in conservation. *Communications Biology*, 3(1), 1–10. <https://doi.org/10.1038/s42003-020-1116-9>
- Quilodrán, C. S., Ruegg, K., Sendell-Price, A. T., Anderson, E. C., Coulson, T., & Clegg, S. M. (2020). The multiple population genetic and demographic routes to islands of genomic divergence. *Methods in Ecology and Evolution*, 11(1), 6–21. <https://doi.org/10.1111/2041-210X.13324>
- Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., Noor, M. A. F., Mehlig, B., & Westram, A. M. (2017). Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, 30(8), 1450–1477. <https://doi.org/10.1111/jeb.13047>
- Roberge, C., Guderley, H., & Bernatchez, L. (2007). Genomewide identification of genes under directional selection: Gene transcription QST scan in diverging atlantic salmon subpopulations. *Genetics*, 177(2), 1011–1022. <https://doi.org/10.1534/genetics.107.073759>
- Roesti, M., Gavrilets, S., Hendry, A. P., Salzburger, W., & Berner, D. (2014). The genomic signature of parallel adaptation from shared genetic variation. *Molecular Ecology*, 23(16), 3944–3956. <https://doi.org/10.1111/mec.12720>
- Rundle, H. D., & Nosil, P. (2005). Ecological speciation. *Ecology Letters*, 8(3), 336–352. <https://doi.org/10.1111/j.1461-0248.2004.00715.x>
- Schlötterer, C. (2003). Hitchhiking mapping - Functional genomics from the population genetics perspective. *Trends in Genetics*, 19(1), 32–38. [https://doi.org/10.1016/S0168-9525\(02\)00012-4](https://doi.org/10.1016/S0168-9525(02)00012-4)

- Schwarz, D., Matta, B. M., Shakir-Botteri, N. L., & McPheron, B. A. (2005). Host shift to an invasive plant triggers rapid animal hybrid speciation. *Nature*, 436(7050), 546–549. <https://doi.org/10.1038/nature03800>
- Sedghifar, A., Brandvain, Y., & Ralph, P. (2016). Beyond clines: lineages and haplotype blocks in hybrid zones. *Molecular Ecology*, 25(11), 2559–2576. <https://doi.org/10.1111/mec.13677>
- Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends in Ecology and Evolution*, 19(4), 198–207. <https://doi.org/10.1016/j.tree.2004.01.003>
- Sendell-Price, A. T., Ruegg, K. C., Anderson, E. C., Quilodrán, C. S., Van Doren, B. M., Underwood, V. L., Coulson, T., & Clegg, S. M. (2020). The Genomic Landscape of Divergence Across the Speciation Continuum in Island-Colonising Silvereyes (*Zosterops lateralis*). *G3: Genes, Genomes, Genetics*, 10(9), 3147–3163. <https://doi.org/10.1534/g3.120.401352>
- Smadbeck, P., & Stumpf, M. P. H. (2016). Coalescent models for developmental biology and the spatio-temporal dynamics of growing tissues. *Journal of the Royal Society Interface*, 13(117). <https://doi.org/10.1098/rsif.2016.0112>
- Stadler, T., Vaughan, T. G., Gavryushkin, A., Guindon, S., Kühnert, D., Leventhal, G. E., & Drummond, A. J. (2015). How well can the exponential-growth coalescent approximate constant-rate birth–death population dynamics? *Proceedings of the Royal Society B: Biological Sciences*, 282(1806). <https://doi.org/10.1098/rspb.2015.0420>
- Supple, M. A., Papa, R., Hines, H. M., McMillan, W. O., & Counterman, B. A. (2015). Divergence with gene flow across a speciation continuum of *Heliconius* butterflies. *BMC Evolutionary Biology*, 15(1). <https://doi.org/10.1186/s12862-015-0486-y>

- Team, R. C. (2017). R Core Team (2017). R: A language and environment for statistical computing. In *R Found. Stat. Comput. Vienna, Austria*. URL <http://www.R-project.org/>, page R Foundation for Statistical Computing.
- Tier, C. (1982). An analysis of neutral-alleles and variable-environment diffusion models. *Journal of Mathematical Biology*, 12(1), 53–71.
- Vijay, N., Bossu, C. M., Poelstra, J. W., Weissensteiner, M. H., Suh, A., Kryukov, A. P., & Wolf, J. B. W. (2016). Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nature Communications*, 7. <https://doi.org/10.1038/ncomms13195>
- Whitlock, M. C. (2008). Evolutionary inference from QST. *Molecular Ecology*, 17(8), 1885–1896. <https://doi.org/10.1111/j.1365-294X.2008.03712.x>
- Wolf, J. B. W., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, 18(2), 87–100. <https://doi.org/10.1038/nrg.2016.133>
- Ye, Z., Wei, W., Pfrender, M. E., & Lynch, M. (2023). Evolutionary Insights from a Large-Scale Survey of Population-Genomic Variation. *Molecular Biology and Evolution*, 40(11). <https://doi.org/10.1093/molbev/msad233>
- Yeaman, S., Aeschbacher, S., & Bürger, R. (2016). The evolution of genomic islands by increased establishment probability of linked alleles. *Molecular Ecology*, 25(11), 2542–2558. <https://doi.org/10.1111/mec.13611>

2. The effect of divergent and parallel selection on the genomic landscape of divergence

Modelling the genomic landscape under selection

Hisham A.A. Ali¹, Tim Coulson¹, Sonya M. Clegg^{1†}, Claudio S. Quilodrán^{1,2†}

¹Edward Grey Institute of Field Ornithology, Department of Biology, University of Oxford, Oxford OX1 3SZ, UK. ²Department of Genetics and Evolution. University of Geneva, CH-1205 Geneva, Switzerland.

† Sonya M. Clegg and Claudio S. Quilodrán are joint senior authors

Correspondence

Hisham A.A. Ali, Department of Biology, University of Oxford, Oxford OX1 3SZ, UK.

Email: hisham.ali@biology.ox.ac.uk

Abstract

While the role of selection in divergence along the speciation continuum is theoretically well understood, defining specific signatures of selection in the genomic landscape of divergence is empirically challenging. Modelling approaches can provide insight into the potential role of selection on the emergence of a heterogeneous genomic landscape of divergence. Here, we extend and apply an individual-based approach that simulates the phenotypic and genotypic distributions of two populations under a variety of selection regimes, genotype-phenotype maps, modes of migration, and genotype-environment interactions. We show that genomic islands of high differentiation and genomic valleys

of similarity may respectively form under divergent and parallel selection between populations. For both types of between-population selection, negative and positive frequency-dependent selection within populations generated genomic islands of higher magnitude and genomic valleys of similarity, respectively. Divergence rates decreased under strong dominance with divergent selection, as well as in models including genotype-environment interactions under parallel selection. For both divergent and parallel selection models, divergence rate was higher under an intermittent migration regime between populations, in contrast to a constant level of migration across generations, despite an equal number of total migrants. We highlight that interpreting a particular evolutionary history from an observed genomic pattern must be done cautiously, as similar patterns may be obtained from different combinations of evolutionary processes. Modelling approaches such as ours provide an opportunity to narrow the potential routes that generate the genomic patterns of specific evolutionary histories.

Keywords: frequency-dependent selection, genomic landscape, genotype-phenotype map, individual-based model, natural selection, speciation

2.1. Introduction

New species can form in a variety of ways (Coyne & Orr, 2004b), but gradual (rather than abrupt) divergence between lineages, culminating in reproductively isolated forms described by the speciation continuum is thought to be frequent (Nosil, Harmon, et al., 2009). Multiple microevolutionary processes (i.e. drift, mutation, migration and different types of selection) can interact to influence progression along this continuum. While

natural selection is considered a primary driver of speciation (Schluter, 2009), reconstructing the roles of, and interactions among, microevolutionary processes as divergence proceeds remains a key goal of evolutionary biology. Genomic signatures of these processes may be observed in the topography of the genomic landscape of divergence: the heterogeneous divergence profile along the genomes of two diverging forms (Gavrilets, 2014; Michel et al., 2010; Wolf & Ellegren, 2017). Previous approaches have provided valuable insights into how microevolutionary processes shape genomic landscapes of divergence (e.g. Andrello & Manel, 2015; Feder, Gejji, et al., 2012; Quilodr  n et al., 2020; Sedghifar et al., 2016). However, these approaches have been restricted to limited demographic parameters and an additive genotype-phenotype map. Additionally, definitive conclusions are difficult because multiple combinations of processes may yield similar genomic profiles (Campbell et al., 2018; Quilodr  n, Ruegg, et al., 2020a; Ravinet et al., 2017; Wolf & Ellegren, 2017). Click or tap here to enter text. Thus, the genomic signatures of combinations of evolutionary histories and processes, interacting with varying genomic architectures, require further investigation.

Two main features of genomic landscapes of divergence that have received particular attention are genomic islands and valleys (Feder, Egan, et al., 2012; Michel et al., 2010; Wolf & Ellegren, 2017). Genomic islands describe regions of the genome with very high levels of divergence, and genomic valleys regions with low levels of divergence, relative to background levels (Ravinet et al., 2017; Roesti et al., 2014b). Different forms of natural selection can result in formation of islands and valleys (Seehausen et al., 2014; Shafer & Wolf, 2013). During the early stages of the speciation continuum, genomic islands are typically formed under divergent selection, where different phenotypes, and hence different alleles, are favoured in each population (Feder et al., 2013; Seehausen et al., 2014). However, the formation of genomic islands is not

restricted to the action of divergent selection, and they can form stochastically at neutral loci (Bay & Ruegg, 2017; Nosil & Feder, 2012; Quilodr  n, Ruegg, et al., 2020a). A genomic valley may form under parallel selection, when the same phenotype is favoured in both populations, and that phenotype is produced by the same allele(s) (Roesti et al., 2014b). However, genomic valleys may also form when similar deleterious mutations are removed in both populations (Cvijovi   et al., 2018). As the speciation continuum progresses, a heterogeneous genomic landscape of divergence may thus emerge between populations influenced by parallel and divergent selection on different aspects of the phenotype (Ravinet et al., 2017; Seehausen et al., 2014).

Frequency dependent selection (FDS) can occur concurrently with parallel and divergent selection during divergence. FDS describes how the frequency of a phenotype influences its fitness within a population, while parallel and divergent selection influence the phenotypic variation between them (Bolnick & Stutz, 2017). By influencing allele frequencies that underlie these phenotypes, FDS can also impact genetic variation between populations. FDS can be positive (PFDS), where the fitness of a phenotype increases as its frequency increases (Endler, 1988; Thompson, 1984), or negative (NFDS) where the fitness of a phenotype increases as its frequency decreases (Clarke & O'Donald, 1962). Both forms have been empirically documented. For instance, PFDS has been observed for hymenopteran black and yellow warning colouration to signal prey unpalatability (Endler, 1988), as well as phlox flower colouration to attract pollinators (Smithson, 2001). NFDS has been observed for predation-driven apostatic selection in the grove snail (*Cepaea nemoralis*) (Allen, 1988); and reproductive mating strategies in the side blotched lizard (*Uta stansburiana*) (Sinervo & Lively, 1996), and ruff (*Philomachus pugnax*) (K  pper et al., 2015). While NFDS acts to maintain genetic diversity (Clarke, 1979; Doebeli & Ispolatov, 2010), PFDS reduces genetic diversity due

to selection against rare phenotypes (Langham, 2004; Mallet & Barton, 1989). However, the potential impact of the interaction of these types of selection with other microevolutionary processes on the genomic landscape of divergence is not well understood (Brisson, 2018; Svensson et al., 2018).

Divergence patterns across the genomic landscape may also be influenced by migration between populations, the genotype-phenotype map and genotype-environment (GxE) interactions. Migration, the movement of alleles between populations, generally acts to hinder genetic divergence (Futuyma, 1987; Morjan & Rieseberg, 2004), though can promote divergence through adaptive introgression, as has been observed in some radiation events (Lamichhaney et al., 2015; Mallet, 2007; Mavárez et al., 2006). The nature of the genotype-phenotype map, and GxE interactions can shape the genomic landscape via effects on genetic variation. The purely genetic contribution to the phenotype can be partitioned into additive, dominance and epistatic variation (Falconer & Mackay, 1996). Where the genotype-phenotype map includes dominance i.e. alleles at a locus interact to produce a phenotype that differs to that expected from the additive effect of those alleles (Falconer & Mackay, 1996), genetic diversity can be maintained when the same phenotype is produced by heterozygous and homozygous dominant loci (Kojima, 1959). GxE interactions, which describe how the phenotype is jointly influenced by the genotype and the environment (Falconer & Mackay, 1996), can also maintain genetic diversity in heterogenous environments (Etges et al., 2007; Mitchell-Olds, 1995).

Several modelling approaches have been developed to simulate the genomic signatures that different microevolutionary processes may produce during divergence. Many of these models are limited to simulating either a single or a few bi-allelic loci, with selection often defined as a single parameter (e.g. Charlesworth et al., 1997; Feder, Gejji, et al., 2012; Feder & Nosil, 2009, 2010; Sedghifar et al., 2016). While considering

selection as a single parameter provides valuable information, it represents a specific stage after a given divergence level is reached (Feder, Gejji, et al., 2012a; Feder & Nosil, 2009, 2010). Coalescence-based models have proved useful, particularly when studying neutral evolution (Lohse, 2017). Coalescence-based models also have the advantage of being able to incorporate large amounts of data, potentially simulating entire chromosomes for large populations, and include selection and migration (Haller et al., 2019; Haller & Messer, 2019; Kelleher et al., 2016). Yet, they are typically less informative for small population sizes affected by selection and non-random reproduction (Currat et al., 2015).

Individual-based models (IBMs), or agent-based models, provide population level insights from the collective behaviour of individuals (DeAngelis & Grimm, 2014; Grimm et al., 2006). The simulated individuals are defined by state variables (e.g. physiological traits or genetic attributes), with fixed population behaviours (e.g. average offspring number and migration rates). IBMs use a forward-in-time approach in contrast to backwards-in-time coalescence-based models (Hoban et al., 2012), allowing the exploration of more complex demographic and genetic dynamics during divergence. In addition, due to the exponential increase in computational power and the wealth of genomic data gathered in recent decades, IBMs have successfully simulated the evolution of populations using large genomic fragments (Currat et al., 2015; Peng & Kimmel, 2005). One challenge of developing IBMs is uncertainty in model structure and its parameters. Natural systems are extremely complex, and so either key aspects of the systems are left out or the system is over-parameterised, and so are of limited use (Grimm et al., 2005). This issue can be mitigated by calibrating the model from observed patterns in the data at different scales (Grimm, 2019). Additionally, the predictions of IBMs are highly dependent on parameterisation, and without precise parameterisation of

the simulated individuals, meaningful interpretations of the data can be challenging (Breckling, 2002). However, this can be mitigated by developing the model to cover a wide parameter space (Grimm, 2019).

A recent application of the IBM approach introduced a flexible fitness function for simulating evolutionary dynamics (Quilodr  n, Ruegg, et al., 2020a). This function allows the linking of evolutionary, ecological, and population demographic parameters in a form that is not possible when selection is summarized as a single parameter. In addition, the framework facilitates simulation of various scenarios of migration, neutral and non-neutral evolution, using non-uniformly distributed loci with varying recombination and mutation rates. We further developed the IBM of Quilodr  n et al., (2020) to model the speciation continuum under both divergent and parallel selection. We introduced four modifications: i) negative and positive frequency dependent selection within populations; ii) genetic dominance rather than a purely additive genotype-phenotype map; iii) genotype-environment interactions; and iv) a modified version of the migration parameter allowing a variable level of migration instead of fixed values across all generations. Our aim is to better understand the formation of heterogeneous genomic landscapes of divergence, including the development of genomic islands and valleys, as well as to provide a more flexible framework for exploring the genomic patterns that may emerge after particular evolutionary histories.

2.2. Methods

2.2.1 Description of the model

The simulations are based on the Genomic Landscape of Divergence Simulations (*'glads'*) IBM developed by Quilodr  n et al., (2020). The *'glads'* framework can be used to explore the role of different microevolutionary processes acting as populations diverge, and via parametrization with empirical data, can also be used to answer questions about specific systems (Sendell-Price et al., 2020) (Fig. 1).

Three hierarchical levels are considered in this framework: genotype, phenotype and demographic parameters. Individuals are assigned a sex and genetic identity. The genotype-phenotype map describes the relationship between different genetic states and the phenotype they produce, by considering different genetic interactions such as epistasis and dominance (Brun-Usan et al., 2022). Here, a genotype-phenotype map defines the expected phenotype from the simulated genotypes, which along with the environment, defines the realized phenotype of individuals. Through the implementation of a fitness function, fitness is computed for each realized phenotype.

Individuals are diploid and are genetically defined by a two-dimensional array that represents homologous chromosomes. Each position in the array represents a locus. With multiple individuals being defined in this manner, we can observe genomic variation at a population level at each locus. For an additive genotype-phenotype map, two integers at each locus influencing a given phenotype determine the locus's breeding values (b_v) (i.e., the homologous alleles in a diploid system), each of which can take the form of any of the simulated alleles e.g. alleles 1 to 20. An individual's phenotype (z) is computed as the sum of all the breeding values contributing to that phenotype (n_a), along with a

stochastic environmental component (ε_{env}), with a defined variance (σ_{env}) and mean (0) (equation 1).

$$z = \sum_{v=1}^{n_a} b_v + \varepsilon_{env}(0, \sigma_{env}) \quad (1)$$

The fitness (ω) of an individual is a modified gaussian function of the phenotype (equation 2). The following population level parameters are introduced in the first part of equation 2: maximum number of offspring (b_0); phenotypic optimum (b_1); and variance of the gaussian distribution of the phenotype (b_2). b_0 represents the maximum fitness of individuals in the population, while b_1 represents the value of z that corresponds with b_0 . In the second part of equation 2, a density-dependent component is introduced to prevent exponential growth (b_3) of the population size (N). In the third part of this equation, a stochastic demographic component (ε_{dem}) with a predefined variance (σ_{dem}) and mean (0) is introduced.

$$\omega = b_0 e^{-\frac{1}{2} \left(\frac{z - b_1 n_a}{b_2 n_a} \right)^2} - b_3 N + \varepsilon_{dem}(0, \sigma_{dem}) \quad (2)$$

The input of each simulation run is composed of different populations, containing multiple genetically and sexually defined individuals, in which population level parameters are applied. Divergent selection can be simulated with different phenotypic optima (b_1) between populations, while parallel selection can be simulated with equal b_1 values. The simulation progresses by iterating discrete steps, representing generations, as found in the Wright-Fisher model. One effect of simulating non-overlapping generations is that the simulated effect of drift is weaker, with estimates from simulations with weak

selection showing that the effect of drift is twice as strong with overlapping generations compared to non-overlapping generation (Rogers & Prugel-Bennett, 2000). In each generation, individuals mate and generate offspring, where the probability of a male siring offspring, and the number of offspring produced by a female depends on their respective fitness. The offspring inherit half of the genome from each parent, but differ from them by specific rules defining recombination and mutation rates. The recombination parameter gives the probability of a crossover occurring; and is implemented as either an average Poisson probability over all loci or as a fixed probability between adjacent loci that considers a recombination map (see Quilodrán et al., 2020). The mutation parameter is defined by a binomial distribution computing the expected number of mutations per site, per generation. This parameter is restricted to biallelic genetic structures, and mutations at selected loci can affect the additive phenotype; consequently, shifting the phenotype towards or away from the phenotypic optimum (adaptive and maladaptive respectively). The offspring reach sexual maturity and have the potential to reproduce in the next generation. Finally, the migration rate parameter defines the probability of sexually mature individuals moving between populations. By applying these parameters over the course of iterative simulations, a genomic landscape of divergence can be observed between populations. By changing the parameters, the effects of different evolutionary mechanisms, such as selection and drift, can be ascertained. This model is implemented in *R* using the ‘*glads*’ package (Quilodrán, Ruegg, et al., 2020a).

2.2.2 Frequency dependent selection

The first modification to the model fitness function (equations 1 and 2) is a framework for frequency dependent selection (FDS). FDS was applied to the phenotype derived in equation 1. The range of phenotypes (z) for each population were partitioned to a pre-

defined number of bins of equal size (K). Individuals (i) were subsequently assigned to a bins (K_j), based on their phenotypic trait value (equation 3).

$$K_j = \left\lceil \frac{z_i - z_{min}}{z_{max} - z_{min}} K \right\rceil \quad (3)$$

FDS was applied to the frequency of individuals within each bin (fK_j), which was normalized by the bin with the highest frequency (fK_{max}). This facilitates simulation of both positive and negative FDS. Under positive FDS, fitness increases as the frequency of individuals within the bin increases. Conversely, under negative FDS, fitness decreases as the bin frequency increases. The strength of FDS was implemented by the parameter (F) that denotes the maximum increase in offspring number due to FDS. The product of parameter F and the normalized bin frequencies ($\frac{fK_j}{fK_{max}}$) represents the fitness change due to FDS (ω'_i), for an individual belonging to a given phenotype bin. In positive FDS, the product is added to the fitness, from equation 2, for each individual (equation 4).

$$\omega'_i = \omega_i + \left(\frac{fK_j}{fK_{max}} F \right) \quad (4)$$

In negative FDS, an inverse relationship between fitness and the bin frequency is computed. This is calculated by subtracting the product of F and the normalized bin frequency ($\frac{fK_j}{fK_{max}}$) from F , as shown in equation 5.

$$\omega'_i = \omega_i + \left(F - \left(\frac{fK_j}{fK_{max}} F \right) \right) \quad (5)$$

2.2.3 Genetic dominance

The second modification made to the model was the addition of dominance to the genotype-phenotype map across all loci. Our genotype-phenotype map considers k alleles per locus influencing a given phenotype, denoted in the IBM array of breeding values (b_v) as 1 to k . We provide an option to define the pattern of the dominance hierarchy in alleles. This is simulated by providing a list of length k numerical values, with each value representing a different allele. These values indicate the position in the dominance hierarchy for the respective allele. Note an equivalent position, or co-dominance, between alleles can be achieved by assigning the alleles equal values in the hierarchy list. Co-dominance was used in the original application of '*glads*' (Quilodrán, Ruegg, et al., 2020a), and was an additive only model. We simulated three different strengths of dominance: strong, medium and an additive model (co-dominance). Under the additive model, each allele in a heterozygote contributed equally to the phenotype, and had an equal chance of being inherited. Therefore, under this model the hierarchy position for all k alleles are equivalent (Fig. S1a). Under the other two models, the chance of inheritance for each allele in a heterozygote was also equal. However, under the strong dominance model, only a single 'dominant' allele would contribute to the phenotype, with a contribution the same as the homozygote for that dominant allele. In equation 1, this corresponds to only a single allele contributing to the b_v for each locus (i.e. the allele with the highest value). In our simulation 20 alleles were used, and the hierarchy for their dominance were the values 1 to 20, ensuring a dominant allele would be present in any heterozygous combination (Fig. S1c). The medium dominance model was the same, except that some heterozygote combinations influenced the phenotype additively instead. Therefore, for equation 1 under this model, either a single or both alleles will contribute

b_v for the locus. To achieve this medium strength dominance, our dominance hierarchy followed a bimodal pattern (Fig. S1b).

The three dominance strengths we simulated are found in natural systems (Billiard et al., 2021). For example, in Brassicaceae and the distantly related *Ipomoea* (Convolvulaceae) and *Senecio* (Asteraceae), linear dominance hierarchies, analogous to our model of strong dominance, have been found for the SP11/SCR gene that controls self-incompatibility (Brennan et al., 2002; Fujii & Takayama, 2018; Kowiyama et al., 1994). Non-linear hierarchies, analogous to our model of medium dominance, include lactose tolerance and sex specific dominance (Billiard et al., 2021; Fry, 2010; Tishkoff et al., 2007). Examples of co-dominance (additive system) include A and B blood type antigens; hemoglobin producing HBB gene, mutations of which cause beta thalassemia and sickle-cell disease; and gametophytic self-incompatibility in the S gene of a number of *Citrus* species (J. H. Kim et al., 2011; Xia, 2013).

2.2.4 Genotype-environment interaction

Genotype-environment (GxE) interactions were introduced to the model's genotype-phenotype map function by modifying equation 1. The first and second components of this linear equation represent the genetic and the environmental effects, respectively. A third component was included in this equation to compute the product of the genotype and the environmental effects, in which parameter (G) represents the strength of this interaction. This product of both components allows us to compute the realized phenotype of each individual by considering the interaction between the genotype and the environment, as shown in equation 6. By incorporating the interaction term, different genotypes produce different phenotypes in response to a particular environment.

$$z = \sum_{v=1}^{n_a} bv + \varepsilon_{env}(0, \sigma_{env}) + (G * \sum_{v=1}^{n_a} bv * \varepsilon_{env}(0, \sigma_{env})) \quad (6)$$

2.2.5 Migration

We introduced a parameter that allows the specification of changing migration rates across generation. We explored two scenarios of migration: i) intermittent migration, in which the migration rate (m_{ij}) oscillates between 0 and 0.01 in incremental steps of 0.001; and ii) constant migration rate (m_{ij}) through time, which represents the original approach included in ‘*glads*’ (see Quilodr  n et al., 2020) (see Fig. S2). Note that for tractability, we only consider migration as a homogeniser that hinders divergence, due to gene flow, and do not consider its potential role in promoting divergence through adaptive introgression.

2.2.6 Initialisation of the model

We parameterised the model following Quilodr  n et al. (2020) to allow comparisons with the original implementation (Table 1). These simulations considered two diploid populations of 400 individuals, with genetic identities composed of 300 loci and 20 alleles per locus. A fraction of the simulated genome was assumed to influence the phenotype additively (i.e., an additive genotype-phenotype map composed of 50 loci). The remaining loci, that were not influencing the phenotype, were considered to be neutral. Because the original implementation focused on patterns at early stages of divergence, our models did not include the effect of new mutations. A recombination map was included, that defined the recombination probability between adjacent loci (ρ), to explore the effect of unlinked ($\rho = 0.5$) and strongly-linked loci ($\rho = 0.0001$). Henceforth, we refer to this original implementation as the base model.

Whilst our approach is a general model building on the parameters used by Quilodr  n et al. (2020), our simulations are inspired by avian systems. Most landbird species (i.e. bird species that live primarily on or over land) produce two to three eggs per clutch, with relatively few producing 7 or more per clutch (Jetz et al., 2008). While our maximum fitness parameter (b_0) of six (Table 1) could be high for a landbird, it is representative of the range of probable values. The size of the FDS percentage corresponds with the increase in fitness of individuals. We therefore set the FDS percentage to 2% FD (Table 1) to mitigate exacerbating an already high maximum clutch size, but also at a percentage that appreciably influences the population. FDS acts upon a phenotype, and due to the additive model containing many loci, a large number of different phenotypes (z) could be generated in the population. We captured this variability by defining 10 phenotypic bins (K) (Table 1). The model is flexible and when modelling natural systems, bin number and strength can be parametrized accordingly.

For our simulations of GxE interaction, we used values from 5 to 25 for the strength parameter (G) (Table 1). This range of values was derived using the partitioning of zebra finch phenotypic variance from Woodgate et al. (2014). The genetic (CV_A), environmental (CV_{EC}) and GxE (CV_{GE}) coefficients of variance were used in the equation $G = CV_{GE} / CV_A \times CV_{GE}$ to derive G for equation 6. In the case of migration, we simulated two patterns: intermittent and constant migration. Under the intermittent model, migration levels varied in a bimodal pattern with a period of cessation of migration in between the peaks of migration (Fig. S2). Under the constant migration model, a fixed migration rate $m_{ij} = 0.006$ was used, which was the average migrate rate of the intermittent model. This value was used to facilitate comparisons of the two models, by simulating an equivalent number of migrants at the end of each simulation. Finally, divergent selection was simulated by assigning the populations distinct phenotypic optima, while parallel

selection was simulated by assigning the populations the same phenotypic optima (b_0 , Table 1). Divergent and parallel selection were parametrised in the same way as implemented in Quilodr  n et al. (2020).

2.2.7 Evaluation of divergence

Genetic divergence between simulated populations was quantified with the fixation index (F_{ST}), the ratio of pairwise differences between the genotypes of two populations, relative to the variation within populations (Weir & Cockerham, 1984). F_{ST} ranges from 0 to 1, where a value of 1 indicates fixation of alternative alleles in each population. F_{ST} was computed every 100 generations across the total of 2000 simulated generations, using the ‘*pegas*’ package (Paradis, 2010). Simulations and analyses were carried out using R (version 3.61; R Core Team, 2019).

A hidden Markov model (HMM) was applied to the F_{ST} output to differentiate among effects of different types of selection included in our simulations (i.e., parallel or divergent selection with frequency dependent selection). We simulated four different types of loci (i.e., selected-linked, selected-unlinked, neutral-linked, and neutral-unlinked), selecting an equal number of randomly chosen loci from each type in order to standardize the HMM. The HMM was used to identify three hidden states: i) genomic islands; ii) genomic valleys; and iii) background levels of differentiation. The HMM analysis was performed following Marques et al. (2016) using the ‘HiddenMarkov’ R package (Harte, 2017). The Baum-Welch algorithm was applied using 1000 randomly chosen initial parameter values to find the parameter estimates with the highest likelihood. The Viterbi algorithm used these parameters to find the most likely sequence of states across the simulated loci. The significance of the assigned states was assessed

by randomly permuting islands and valleys 10,000 times across the simulated loci (Quilodr  n, Ruegg, et al., 2020a).

2.2.8 Simulations

Our simulations were designed to elucidate how microevolutionary processes may influence the emergence of a heterogeneous genomic landscape of divergence. The parameters used for the simulations are listed in Table 1. The simulations explored how 1) frequency dependent selection, 2) dominance, 3) GxE interaction, and 4) migration can influence the genomic landscape, alongside the effect of divergent selection, parallel selection, and drift (Table 2). While we analyse each effect separately, the concurrent effect of these processes may be simulated using the ‘*glads*’ model. The code for the modified *glads* functions can be found in supporting information (Appendix S1).

All simulations were run for a maximum of 2000 generations. The fixation index (F_{ST}) was measured every 100 generations to characterize the formation of the genomic landscape of divergence. We also measured the phenotypic space, measured as the range of a quantitative phenotype occupied by a population. When the effect of the microevolutionary process being investigated was not easily interpretable, the phenotypic space provided additional data on how a particular genomic landscape formed. Each simulation was repeated 100 times, with the individuals for each iteration varying in genetic identity at the beginning (generation 0). This was implemented by using a different random number generator in *R* for each simulation (*set.seed* function). These simulations were performed at the University of Geneva on the "Yggdrasil" HPC cluster and on the Oxford ARC clusters.

2.3. Results

2.3.1 Frequency-dependent selection

Our first set of simulations explored the effect of frequency dependent selection on the genomic landscape of divergence between two populations that have evolved independently over 1000 generations (i.e., without migration). In the base model (a purely additive genotype-phenotype map without FDS), the genomic landscape under divergent selection shows the formation of a genomic island at selected loci (Fig. 2a) while parallel selection does not produce any feature that can be distinguished from neutral background divergence (Fig. 2b). Implementation of PFDS produces a genomic island under both divergent and parallel selection regimes, while NFDS produces a genomic valley under both regimes (Fig. 2a,b). The genomic island formed under PFDS is not influenced by linkage to any great extent. The entire selected region approach fixation regardless of position of linked or unlinked loci (Fig. 2a,b), with little variation in this outcome over all simulations (Fig. 3a,b). This contrasts with both the baseline model and NFDS where the island or valley signature is amplified at linked loci (Fig. 2a,b; Fig. 3a,b). In general, a high degree of variation in F_{ST} was seen across the 100 simulations for different combinations of selected/neutral and linked/unlinked loci (Fig. 3a,b).

While islands or valleys were frequently observed on selected linked loci, significant islands and valleys (as identified by the hidden Markov model) were also observed elsewhere (Fig. 4). For instance, the majority of valleys observed for both divergent and parallel selection under PFDS occurred at neutral loci (both linked and unlinked), and the majority of islands formed under NFDS also occurred at neutral loci (Fig. 4). Under parallel selection in the baseline additive model, the number of islands

and valleys for different loci types (selected/neutral and linked/unlinked loci combinations) were similar (Fig. 4b). However, under divergent selection, the majority of the islands and a smaller proportion of valleys were found at selected loci (Fig. 4a). As expected, under PFDS for both types of selection, a higher number of islands on selected unlinked loci were observed compared to the base additive model (Fig. 4c,f). The expected contrast was observed with NFDS, in which an excess of valleys and very few islands were observed on linked selected loci for both types of selection (Fig. 4b,e).

A comparison of the genomic landscape of divergence at 1000 generations (above results) with that at 100 generations and 2000 generations showed the dynamics of genomic landscape features develop over time (Fig. 5). At 100 generations, incipient island formation was already evident, and for the base model under divergent selection, and PFDS under both divergent and parallel selection, these islands persisted at 2000 generations, albeit at a smaller amplitude. Valley formation occurred later in the divergence process when NFDS was applied, with an island rather than a valley evident in early stage divergence (100 generations) (Fig. 5). The depth of valleys increased between 1000 and 2000 generations as surrounding neutral divergence increased.

The distribution of phenotypes realized by individuals at a range of generation time points under divergent and parallel selection, and with and without FDS is shown in Fig. 6. At early stages of divergence (<100 generations), the distribution of phenotypes was similar for all models. The implementation of divergent selection with the base model, or with PFDS, resulted in two disjunct phenotypic groups by 500 generations, each occupying a relatively narrow range of values. Under divergent selection with NFDS, the two populations overlapped in phenotypic space after 2000 generations despite the divergence of the phenotypic space of the populations. Under parallel selection, phenotypes in the two populations did not diverge, but the range of phenotypes

is narrowed when applied to the base model, or with PFDS, compared to a wider range for NFDS. In short, the models showed the expected changes in between- and within-population variation under different population selection regimes.

These results highlight that resulting levels of divergence were highly variable. For each independent simulation, the starting genomic composition varied, and under these varying conditions, with the same parameters, similar values of divergence could be obtained across loci irrespective of linkage and selection. This is shown by the HMM, where the formation of islands and valleys potentially occur across all loci types (Fig. 4).

2.3.2 Dominance

The next set of models modified the genotype-phenotype map to include the effects of dominance rather than alleles acting in a purely additive manner. After 1000 generations of independent evolution, and application of a divergent selection scenario, an island of divergence was evident when moderate dominance was applied (similar to, though slightly higher in amplitude, a purely additive model), however application of strong dominance inhibited island formation altogether (Fig. 2c). Under parallel selection, pronounced islands of divergence were not seen for either moderate or strong dominance, as was the case for the base additive model (Fig. 3d).

The development of the genomic landscape of divergence over generations is shown in Fig. S3. When dominance was strong, divergence levels were comparable for selected and neutral, and linked and unlinked loci from the earliest stages of divergence (100 generations) through to the latest stages (2000 generations), with only a slightly elevated selected region under divergent selection. The genomic island that formed under moderate dominance level and divergent selection was observed from the earliest stages.

A transient, low amplitude genomic island developed in the early stages under parallel selection and moderate dominance.

2.3.3 Genotype x Environment interaction

The addition of GxE interactions of varying strengths ($G = 5$ to 25 , see eq.6) did not change the shape of the landscape of divergence to any large degree under divergent selection (Fig. 2e). However, inclusion of GxE interactions depressed divergence across the genome compared to the base additive model when parallel selection was applied (Fig. 2f). This difference from the additive model was maintained from the early to latest stage of divergence (Fig. S4). Despite the strength of the applied GxE interaction, under varying initial genetic diversity, similar values of divergence could be obtained across loci irrespective of linkage and selection (Fig. 3e,f).

2.3.4 Migration

The addition of migration, whether acting in a constant or intermittent manner, slowed divergence overall compared to a model with no migration, but the main features of the profile of the genomic landscape were maintained: the formation of an island under divergent selection, which was more prominent at linked loci; and divergence levels equivalent to neutral loci under parallel selection (Fig. 2g,h). However, delivery of the same amount of migration intermittently rather than constantly produced higher levels of divergence. This was more pronounced under a parallel selection regime, where the difference emerged early in the process (100 generations) and was maintained until the end of the simulations (2000 generations) (Fig. S5). The variation around the average values obtained from the 100 independent simulations after 1000 generations showed

that, under divergent selection for both migration models, selected loci had higher levels of divergence than neutral loci (Fig. 3g,h).

2.4. Discussion

Our individual-based modelling approach has revealed the shapes of the genomic landscape of divergence when varying (1) the type of selection acting within and between populations, and (2) the form of the genotype-phenotype map. Genomic islands or valleys may form irrespective of whether between-population selection is divergent or parallel when positive and negative frequency dependent selection also operate within each population, and can also form in the absence of selection (Fig. 2a,b). Furthermore, we show that a consequence of dominance in the genotype-phenotype map is the erosion of genomic islands under divergent selection. The level of divergence at both neutral and selected loci were highly variable in our simulations, and the formation of islands or valleys was not always guaranteed. This variable output is known to be influenced by a range of factors (Ravinet et al., 2017; Yeaman et al., 2016), including the initial genetic diversity of the ancestral population and drift (Quilodr  n, Ruegg, et al., 2020a; Yeaman et al., 2016). Our findings confirm that the genotype-phenotype map, gene flow pattern, and selection regime affect the development of a heterogeneous genomic landscape of divergence (Quilodr  n, Ruegg, et al., 2020a). We also show they affect the maintenance of genetic diversity under certain conditions: negative frequency dependent selection and dominance.

2.4.1 A heterogenous genomic landscape of divergence under divergent and parallel selection is influenced by within-population frequency dependent selection.

Frequency dependent selection within populations (PFDS and NFDS) influenced the genomic landscape over time, irrespective of the type of selection between populations (i.e. divergent and parallel). Both PFDS and NFDS reflected our expectation of generating islands and valleys of divergence, respectively. High linkage influenced the shape of the genomic landscape, with selection acting effectively on entire haplotype blocks that then rapidly increase in frequency. This corroborates empirical results where strong linkage typically associated with genomic inversions or proximity to centromeres, is often associated with heterogenous genomic landscapes (Huang et al., 2020; Puig Giribets et al., 2019; Sodeland et al., 2016; Tepolt & Palumbi, 2020).

PFDS resulted in a genomic island that was not evident under parallel selection alone, and was also taller under the effects of divergent selection. PFDS is known to accelerate the fixation of the most frequently occurring allele at a locus under selection (Gordon et al., 2015), which results in a rapid reduction of within population phenotypic variation in our simulations. While this increased the difference between quantitative phenotypes of populations under divergent selection, it decreased the difference under parallel selection, with both populations reaching similar phenotypes. Because there are multiple genetic combinations that result in a given quantitative phenotype, small differences in frequency of the most common allele in each population could be quickly amplified by PFDS, resulting in the development of genomic islands over time. Experimental and empirical investigations have shown the proliferation of monomorphism in species influenced by PFDS (Lindström et al., 2001; Mallet & Barton, 1989; Nokelainen et al., 2014), but also polymorphism (Chouteau et al., 2016; Ogilvie et

al., 2021; Rönkä et al., 2020). Both are compatible with our expectation of PFDS under different types of selective pressure between populations.

NFDS as a type of balancing selection is expected to maintain ancestral polymorphism (Brisson, 2018). Genetic diversity within populations negatively correlates with divergence between populations, which may in turn exacerbate valley formation (Wolf & Ellegren, 2017). In our simulations, this is illustrated by the formation of genomic valleys irrespective of the form of selection between populations. The influence of NFDS maintained the within population phenotypic variation and consequently decreased the distance between the phenotypic space of the populations. This finding is in line with empirical observations made under allopatric conditions of morphological overlap between populations under NFDS (Franks & Oxford, 2017). However, under NFDS, the genomic landscape feature for linked loci often transitioned from a genomic island seen at early stages of divergence (100 generations), to a valley by the later stages (1000 generations). This can be explained by the stochastic generation of different additive routes towards a common phenotype, forming idiosyncratic differences between the populations, resulting in genomic islands in some early stage simulations (Goldberg et al., 2020). However, as divergence progresses, the cumulative effect of NFDS generates large within-population genotypic and phenotypic variation, resulting in genomic valleys at later stages for both parallel and divergent selection between populations.

2.4.2 Maintenance of polymorphism under divergent selection limits island formation

The incorporation of polymorphism-maintaining processes such as dominance and NFDS (Brisson, 2018; Moulherat et al., 2017) shape the genomic landscape of divergence. Under strong dominance and divergent selection, genomic island formation was inhibited

as recessive alleles were maintained in the heterozygous form. Other mechanisms not explored here by which dominance could maintain polymorphism include situations where dominance relationships among alleles reverse, for example where dominance is sex-dependent or non-linear relationships exist between fitness and gene activity (Connallon & Chenoweth, 2019; Nabutanyi & Wittmann, 2021). When the average strength of dominance decreases (under the medium dominance model), dominant-recessive combinations of alleles form more rarely, polymorphism is therefore maintained more rarely, allowing formation of islands of divergence as seen in our simulations. NFDS is even more effective at maintaining polymorphism than dominance because the latter is not a selective process (Asmussen et al., 2004), resulting in valley rather than island formation at selected loci. Overall, the incorporation of polymorphism maintaining processes could help reconcile differences between empirical and simulated levels of divergence. For example, in the silveryeyes species complex (*Zosterops lateralis*) simulated levels of divergence exceed empirical observations (Sendell-Price et al., 2020).

2.4.3 Genotype-Environment interaction reduces divergence of selected and neutral loci under parallel selection

The effects of GxE interactions on the genomic landscape were our most surprising results; particularly, the lack of effect GxE interaction had on the landscape under divergent selection compared to parallel selection. In heterogeneous environments, GxE interactions can favour different genotypes in different environments, maintaining allelic diversity (Côté & Simons, 2020; Falconer & Mackay, 1996; Plesnar-Bielak et al., 2018). In our simulations, the phenotype was an additive combination of both the stochastic environmental component (varied between individuals in a population) and the additive genotypic component. As a result, different additive genotypes complemented different

environmental components within a population (e.g. a given additive genotype may form a fitter phenotype with a particular environmental value and a less fit phenotype with a different environmental condition) (Falconer & Mackay, 1996). Furthermore, since the genotype – phenotype function in equation 6 additively includes the product of the genotype and the environment, higher values of GxE also increase the phenotypic variance of the population. Our simulations reflect the increased phenotypic variance, as a consequence of the implementation of GxE in equation 6. Under parallel selection, a consequence of the increased genetic variance was that more often common allelic combinations persisted in the simulations, decreasing F_{ST} . However, under divergent selection, different phenotypes were selectively favoured between both populations, in which alternate allelic combinations (generating alternate phenotypes) were favoured. In this last condition, the frequency of common allelic combinations between the two populations was not affected by the increase in phenotypic variance (as a consequence of GxE interaction), and so F_{ST} did not decrease.

2.4.4 Intermittent gene flow slows the loss of genomic islands over time

Genomic islands of divergence are expected to disappear over time with the accumulation of genome-wide differences between populations (Nosil & Feder, 2012; Wu & Ting, 2004). However, the temporal dynamic of this process can be moderated by gene flow through migration, in which some gene flow slows the process of genomic island erosion, while a high level can rapidly erase them (Quilodr  n, Ruegg, et al., 2020a). We show that the gene flow pattern (intermittent or continuous) influences the overall level of divergence, with intermittent gene flow resulting in a higher level of divergence despite an equal number of migrants, via a slowed loss of genomic islands over time. This emphasises the importance of considering different patterns of gene flow when aiming to

depict the historical divergence of any biological system (Kirkpatrick & Ravigné, 2002; Rundle & Nosil, 2005)

2.5. Conclusion

The interpretation of past evolutionary histories based on the observation of given genomic patterns have to be made with caution (e.g. Funk et al., 2021 and Ottenburghs et al., 2017), as the emergence of specific patterns is not ensured by a unique set of circumstances. While divergent selection with ongoing gene flow was a prominent explanation for genomic islands of divergence (Feder et al., 2013), more recent studies have shown that they may emerge from a variety of evolutionary routes in the absence of gene flow (Ravinet et al., 2017). Even under divergent selection and strong linkage, two elements that often show patterns of genomic islands when concurrently influencing genomic landscapes (Nam et al., 2020), we have shown that islands could disappear or even turn out to be valleys of similarity when adding other layers of complexity to the system, such as strong dominance and frequency dependent selection. Computational approaches that concurrently analyse theoretical expectations with summary statistics based on empirical data (e.g., Approximate Bayesian computational methods and machine learning methods) provide an opportunity to narrow the range of potential drivers that produce a particular genomic landscape of divergence under defined evolutionary histories. The modelling tool used in our approach will allow further exploration of genomic landscape dynamics under the nuanced and complex conditions that characterize biological systems. Empirical data can be fitted to this flexible model to simulate the evolution of natural populations (e.g. Sendell-Price et al., 2020). In doing so,

our individual-based approach may help to provide powerful insights into the effects of different microevolutionary processes on particular natural systems.

Acknowledgements

The authors would like to acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility (<http://dx.doi.org/10.5281/zenodo.22558>) and the University of Geneva “Yggdrasil” HPC cluster. CSQ acknowledges support from the Swiss National Science Foundation (P5R5PB_203169).

Author Contributions

H.A.A.A., T.C., S.M.C., and C.S.Q. conceived the original idea for the study. H.A.A.A coded and performed the simulations in R. H.A.A.A wrote the first draft of the manuscript and all authors provided comments and edits on manuscript versions.

Data Accessibility Statement

All information needed to repeat our simulations are described in Table 1. They can be reproduced by using the R package ‘glads’, which is available on GitHub and deposited in Zenodo: <https://doi.org/10.5281/zenodo.3516224>. The custom functions used in the manuscript are available in the supporting information Appendix S1.

Supporting Information

Additional supporting information is available online for this article.

Fig. S1-S5 contains supporting figures for this article

Appendix S1 contains the custom code used in this article with examples.

2.6. References

- Allen, J. A. (1988). Frequency-dependent selection by predators. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 319(1196), 485–503.
- Andrello, M., & Manel, S. (2015). MetaPopGen: An R package to simulate population genetics in large size metapopulations. *Molecular Ecology Resources*, 15(5), 1153–1162. <https://doi.org/10.1111/1755-0998.12371>
- Asmussen, M. A., Cartwright, R. A., & Spencer, H. G. (2004). Frequency-dependent selection with dominance: A window onto the behavior of the mean fitness. *Genetics*, 167(1), 499–512. <https://doi.org/10.1534/genetics.167.1.499>
- Bay, R. A., & Ruegg, K. (2017). Genomic islands of divergence or opportunities for introgression? *Proceedings of the Royal Society B: Biological Sciences*, 284(1850). <https://doi.org/10.1098/rspb.2016.2414>
- Billiard, S., Castric, V., & Llaurens, V. (2021). The integrative biology of genetic dominance. *Biological Reviews*, 96(6), 2925–2942. <https://doi.org/10.1111/brv.12786>
- Bolnick, D. I., & Stutz, W. E. (2017). Frequency dependence limits divergent evolution by favouring rare immigrants over residents. *Nature*, 546(7657), 285–288. <https://doi.org/10.1038/nature22351>
- Breckling, B. (2002). Individual-based modelling: potentials and limitations. In *TheScientificWorldJournal* (Vol. 2, pp. 1044–1062). <https://doi.org/10.1100/tsw.2002.179>
- Brennan, A. C., Harris, S. A., Tabah, D. A., & Hiscock, S. J. (2002). The population genetics of sporophytic self-incompatibility in *Senecio squalidus* L. (Asteraceae) I: S allele

- diversity in a natural population. *Heredity*, 89(6), 430–438.
<https://doi.org/10.1038/sj.hdy.6800159>
- Brisson, D. (2018). Negative frequency-dependent selection is frequently confounding. *Frontiers in Ecology and Evolution*, 6(FEB), 1–9.
<https://doi.org/10.3389/fevo.2018.00010>
- Brun-Usan, M., Zimm, R., & Uller, T. (2022). Beyond genotype-phenotype maps: Toward a phenotype-centered perspective on evolution. *BioEssays*, 44(9).
<https://doi.org/10.1002/bies.202100225>
- Campbell, C. R., Poelstra, J. W., & Yoder, A. D. (2018). What is Speciation Genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species. *Biological Journal of the Linnean Society*, 124(4), 561–583.
<https://doi.org/10.1093/biolinnean/bly063>
- 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.
<https://doi.org/10.1017/S0016672397002954>
- Chouteau, M., Arias, M., & Joron, M. (2016). Warning signals are under positive frequency dependent selection in nature. *Proceedings of the National Academy of Sciences of the United States of America*, 113(8), 2164–2169. <https://doi.org/10.1073/pnas.1519216113>
- Clarke, B. C. (1979). The Evolution of Genetic Diversity. *Proceedings of the Royal Society B*, 205(1161), 453–474. <https://doi.org/https://doi.org/10.1098/rspb.1979.0079>
- Clarke, B. C., & O'Donald, P. (1962). Frequency-dependent selection. *Heredity*, 19(2), 201–206. <https://doi.org/https://doi.org/10.1038/hdy.1964.25>

- Connallon, T., & Chenoweth, S. F. (2019). Dominance reversals and the maintenance of genetic variation for fitness. *PLoS Biology*, 17(1), 1–11. <https://doi.org/10.1371/journal.pbio.3000118>
- Côté, K., & Simons, A. M. (2020). Genotype-environment interaction and the maintenance of genetic variation: An empirical study of *Lobelia inflata* (Campanulaceae). *Royal Society Open Science*, 7(3). <https://doi.org/10.1098/rsos.191720>
- Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sinauer. <https://doi.org/10.1126/science.1187468>
- Curat, M., Gerbault, P., Di, D., Nunes, J. M., & Sanchez-Mazas, A. (2015). Forward-in-time, spatially explicit modeling software to simulate genetic lineages under selection. *Evolutionary Bioinformatics*, 11, 27–39. <https://doi.org/10.4137/EBO.S33488>
- Cvijović, I., Good, B., & Desai, M. (2018). The effect of strong purifying selection on genetic diversity. *The Effect of Strong Purifying Selection on Genetic Diversity*, 209(August), 1235–1278. <https://doi.org/10.1101/211557>
- DeAngelis, D. L., & Grimm, V. (2014). Individual-based models in ecology after four decades. *F1000Prime Reports*, 6(June). <https://doi.org/10.12703/P6-39>
- Doebeli, M., & Ispolatov, I. (2010). Complexity and diversity. *Science*, 328(5977), 494–497. <https://doi.org/10.1126/science.1187468>
- Endler, J. A. (1988). Frequency-dependent predation, crypsis and aposematic coloration. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 319(1196), 505–523. <https://doi.org/https://doi.org/10.1098/rstb.1988.0062>
- Etges, W. J., De Oliveira, C. C., Gragg, E., Ortíz-Barrientos, D., Noor, M. A. F., & Ritchie, M. G. (2007). Genetics of incipient speciation in *Drosophila mojavensis*. I. Male courtship song, mating success, and genotype X environment interactions. *Evolution*, 61(5), 1106–1119. <https://doi.org/10.1111/j.1558-5646.2007.00104.x>

- Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to quantitative genetics* (4th ed.). Longman.
- Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics*, 28(7), 342–350. <https://doi.org/10.1016/j.tig.2012.03.009>
- Feder, J. L., Flaxman, S. M., Egan, S. P., Comeault, A. A., & Nosil, P. (2013). Geographic mode of speciation and genomic divergence. *Annual Review of Ecology, Evolution, and Systematics*, 44(August), 73–97. <https://doi.org/10.1146/annurev-ecolsys-110512-135825>
- Feder, J. L., Gejji, R., Yeaman, S., & Nosil, P. (2012). Establishment of new mutations under divergence and genome hitchhiking. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1587), 461–474. <https://doi.org/10.1098/rstb.2011.0256>
- Feder, J. L., & Nosil, P. (2009). Chromosomal inversions and species differences: When are genes affecting adaptive divergence and reproductive isolation expected to reside within inversions? *Evolution*, 63(12), 3061–3075. <https://doi.org/10.1111/j.1558-5646.2009.00786.x>
- Feder, J. L., & Nosil, P. (2010). The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution: International Journal of Organic Evolution*, 64(6), 1729–1747. <https://doi.org/10.1111/j.1558-5646.2009.00943.x>
- Franks, D. W., & Oxford, G. S. (2017). The co-evolution of anti-predator polymorphisms in sympatric populations. *Biological Journal of the Linnean Society*, 122(4), 729–737. <https://doi.org/10.1093/biolinnean/blx111>
- Fry, J. D. (2010). The genomic location of sexually antagonistic variation: some cautionary comments. *Evolution: International Journal of Organic Evolution*, 64(5), 1510–1516. <https://doi.org/10.1111/j.1558-5646.2009.00898.x>

- Fujii, S., & Takayama, S. (2018). Multilayered dominance hierarchy in plant self-incompatibility. *Plant Reproduction*, 31(1), 15–19. <https://doi.org/10.1007/s00497-017-0319-9>
- Funk, E. R., Spellman, G. M., Winker, K., Withrow, J. J., Ruegg, K. C., Zavaleta, E., & Taylor, S. A. (2021). Phylogenomic Data Reveal Widespread Introgression Across the Range of an Alpine and Arctic Specialist. *Systematic Biology*, 70(3), 527–541. <https://doi.org/10.1093/sysbio/syaa071>
- Futuyma, D. J. (1987). On the Role of Species in Anagenesis. *American Naturalist*, 130(3), 465–473. <https://doi.org/10.1086/284724>
- Gavrilets, S. (2014). Models of speciation: Where are we now? *Journal of Heredity*, 105(S1), 743–755. <https://doi.org/10.1093/jhered/esu045>
- Goldberg, J. K., Lively, C. M., Sternlieb, S. R., Pintel, G., Hare, J. D., Morrissey, M. B., & Delph, L. F. (2020). Herbivore-mediated negative frequency-dependent selection underlies a trichome dimorphism in nature. *Evolution Letters*, 4(1), 83–90. <https://doi.org/10.1002/evl3.157>
- Gordon, S. P., Kokko, H., Rojas, B., Nokelainen, O., & Mappes, J. (2015). Colour polymorphism torn apart by opposing positive frequency-dependent selection, yet maintained in space. *Journal of Animal Ecology*, 84(6), 1555–1564. <https://doi.org/10.1111/1365-2656.12416>
- Grimm, V. (2019). Ecological Models: Individual-Based Models☆. In B. Fath (Ed.), *Encyclopedia of Ecology (Second Edition)* (pp. 65–73). Elsevier. <https://doi.org/https://doi.org/10.1016/B978-0-12-409548-9.11144-3>
- Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S. K., Huse, G., Huth, A., Jepsen, J. U., Jørgensen, C., Mooij, W. M.,

- Müller, B., Pe'er, G., Piou, C., Railsback, S. F., Robbins, A. M., ... DeAngelis, D. L. (2006). A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*, 198(1–2), 115–126. <https://doi.org/10.1016/j.ecolmodel.2006.04.023>
- Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W. M., Railsback, S. F., Thulke, H.-H., Weiner, J., Wiegand, T., & DeAngelis, D. L. (2005). Pattern-oriented modeling of agent-based complex systems: lessons from ecology. *Science*, 310(5750), 987–991.
- Haller, B. C., Galloway, J., Kelleher, J., Messer, P. W., & Ralph, P. L. (2019). Tree-sequence recording in SLiM opens new horizons for forward-time simulation of whole genomes. *Molecular Ecology Resources*, 19(2), 552–566. <https://doi.org/10.1111/1755-0998.12968>
- Haller, B. C., & Messer, P. W. (2019). SLiM 3: Forward Genetic Simulations Beyond the Wright-Fisher Model. *Molecular Biology and Evolution*, 36(3), 632–637. <https://doi.org/10.1093/molbev/msy228>
- Harte, D. (2017). *HiddenMarkov: Hidden Markov Models* (R package version 1.8-11). Statistics Research Associates. <http://www.statsresearch.co.nz/dsh/sslib/>
- Hoban, S., Bertorelle, G., & Gaggiotti, O. E. (2012). Computer simulations: Tools for population and evolutionary genetics. *Nature Reviews Genetics*, 13(2), 110–122. <https://doi.org/10.1038/nrg3130>
- Huang, K., Andrew, R. L., Owens, G. L., Ostevik, K. L., & Rieseberg, L. H. (2020). Multiple chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype. *Molecular Ecology*, 29(14), 2535–2549. <https://doi.org/10.1111/mec.15428>
- Jetz, W., Sekercioglu, C. H., & Böhning-Gaese, K. (2008). The worldwide variation in avian clutch size across species and space. *PLoS Biology*, 6(12), 2650–2657. <https://doi.org/10.1371/journal.pbio.0060303>

- Kelleher, J., Etheridge, A. M., & McVean, G. (2016). Efficient Coalescent Simulation and Genealogical Analysis for Large Sample Sizes. *PLoS Computational Biology*, 12(5), 1–22. <https://doi.org/10.1371/journal.pcbi.1004842>
- Kim, J. H., Mori, T., Wakana, A., Ngo, B. X., Sakai, K., & Kajiwara, K. (2011). Determination of self-incompatible citrus cultivars with S1 and/or S2 alleles by pollination with homozygous S1 seedlings (S1S1 or S2S2) of “Banpeiyu” pummelo. *Journal of the Japanese Society for Horticultural Science*, 80(4), 404–413. <https://doi.org/10.2503/jjshs1.80.404>
- Kirkpatrick, M., & Ravigné, V. (2002). Speciation by natural and sexual selection: Models and experiments. *American Naturalist*, 159(3 SUPPL.). <https://doi.org/10.2307/3078919>
- Kojima, K. (1959). Stable equilibria for the optimum model. *Proceedings of the National Academy of Sciences of the United States of America*, 45(7), 989. <https://doi.org/10.1073/pnas.45.7.989>
- Kowyama, Y., Takahasi, H., Muraoka, K., Tani, T., Hara, K., & Shiotani, I. (1994). Number, frequency & dominance relationships of S-alleles in diploid *Ipomoea trifida*. *Heredity*, 73(3), 275–283. <https://doi.org/10.1038/hdy.1994.134>
- Küpper, C., Stocks, M., Risse, J. E., Dos Remedios, N., Farrell, L. L., McRae, S. B., Morgan, T. C., Karlionova, N., Pinchuk, P., Verkuil, Y. I., Kitaysky, A. S., Wingfield, J. C., Piersma, T., Zeng, K., Slate, J., Blaxter, M., Lank, D. B., & Burke, T. (2015). A supergene determines highly divergent male reproductive morphs in the ruff. *Nature Genetics*, 48(1), 79–83. <https://doi.org/10.1038/ng.3443>
- Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., Promerová, M., Rubin, C. J., Wang, C., Zamani, N., Grant, B. R., Grant, P. R., Webster, M. T., & Andersson, L. (2015). Evolution of Darwin’s finches and their beaks

- revealed by genome sequencing. *Nature*, 518(7539), 371–375.
<https://doi.org/10.1038/nature14181>
- Langham, G. M. (2004). Specialized avian predators repeatedly attack novel color morphs of *Heliconius* butterflies. *Evolution*, 58(12), 2783–2787. <https://doi.org/10.1111/j.0014-3820.2004.tb01629.x>
- Lindström, L., Alatalo, R. V., Lyytinen, A., & Mappes, J. (2001). Strong antiapostatic selection against novel rare aposematic prey. *Proceedings of the National Academy of Sciences of the United States of America*, 98(16), 9181–9184.
<https://doi.org/10.1073/pnas.161071598>
- Lohse, K. (2017). Come on feel the noise – from metaphors to null models. *Journal of Evolutionary Biology*, 30(8), 1506–1508. <https://doi.org/10.1111/jeb.13109>
- Mallet, J. (2007). Hybrid speciation. *Nature*, 446(7133), 279–283.
<https://doi.org/10.1038/nature05706>
- Mallet, J., & Barton, N. H. (1989). Strong natural selection in a warning-color hybrid zone. *Evolution*, 43(2), 421–431. <https://doi.org/10.1111/j.1558-5646.1989.tb04237.x>
- Marques, D. A., Lucek, K., Meier, J. I., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2016). Genomics of Rapid Incipient Speciation in Sympatric Threespine Stickleback. *PLoS Genetics*, 12(2), 1–34. <https://doi.org/10.1371/journal.pgen.1005887>
- Mavárez, J., Salazar, C. A., Bermingham, E., Salcedo, C., Jiggins, C. D., & Linares, M. (2006). Speciation by hybridization in *Heliconius* butterflies. *Nature*, 441(7095), 868–871.
<https://doi.org/10.1038/nature04738>
- Michel, A. P., Sim, S., Powell, T. H. Q., Taylor, M. S., Nosil, P., & Feder, J. L. (2010). Widespread genomic divergence during sympatric speciation. *Proceedings of the National Academy of Sciences of the United States of America*, 107(21), 9724–9729.
<https://doi.org/10.1073/pnas.1000939107>

- Mitchell-Olds, T. (1995). The molecular basis of quantitative genetic variation in natural populations. *Trends in Ecology & Evolution*, 10(8), 324–328. [https://doi.org/10.1016/S0169-5347\(00\)89119-3](https://doi.org/10.1016/S0169-5347(00)89119-3)
- Morjan, C. L., & Rieseberg, L. H. (2004). How species evolve collectively: Implications of gene flow and selection for the spread of advantageous alleles. *Molecular Ecology*, 13(6), 1341–1356. <https://doi.org/10.1111/j.1365-294X.2004.02164.x>
- Moulherat, S., Chaine, A., Mangin, A., Aubret, F., Sinervo, B., & Clobert, J. (2017). The roles of plasticity versus dominance in maintaining polymorphism in mating strategies. *Scientific Reports*, 7(1), 1–10. <https://doi.org/10.1038/s41598-017-15078-1>
- Nabutanyi, P., & Wittmann, M. J. (2021). Models for eco-evolutionary extinction vortices under balancing selection. *American Naturalist*, 197(3), 336–350. <https://doi.org/10.1086/712805>
- Nam, K., Nhim, S., Robin, S., Bretaudeau, A., Nègre, N., & d'Alençon, E. (2020). Positive selection alone is sufficient for whole genome differentiation at the early stage of speciation process in the fall armyworm. *BMC Evolutionary Biology*, 20(1), 1–16. <https://doi.org/10.1186/s12862-020-01715-3>
- Nokelainen, O., Valkonen, J., Lindstedt, C., & Mappes, J. (2014). Changes in predator community structure shifts the efficacy of two warning signals in Arctiid moths. *Journal of Animal Ecology*, 83(3), 598–605. <https://doi.org/10.1111/1365-2656.12169>
- Nosil, P., & Feder, J. L. (2012). Genomic divergence during speciation: Causes and consequences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1587), 332–342. <https://doi.org/10.1098/rstb.2011.0263>
- Nosil, P., Harmon, L. J., & Seehausen, O. (2009). Ecological explanations for (incomplete) speciation. *Trends in Ecology and Evolution*, 24(3), 145–156. <https://doi.org/10.1016/j.tree.2008.10.011>

- Ogilvie, J. G., Van Belleghem, S., Range, R., Papa, R., McMillan, O. W., Chouteau, M., & Counterman, B. A. (2021). Balanced polymorphisms and their divergence in a *Heliconius* butterfly. *Ecology and Evolution*, *11*(24), 18319–18330. <https://doi.org/10.1002/ece3.8423>
- Ottenburghs, J., Megens, H. J., Kraus, R. H. S., Van Hooft, P., Van Wieren, S. E., Crooijmans, R. P. M. A., Ydenberg, R. C., Groenen, M. A. M., & Prins, H. H. T. (2017). A history of hybrids? Genomic patterns of introgression in the True Geese. *BMC Evolutionary Biology*, *17*(1), 1–14. <https://doi.org/10.1186/s12862-017-1048-2>
- Paradis, E. (2010). Pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics*, *26*(3), 419–420. <https://doi.org/10.1093/bioinformatics/btp696>
- Peng, B., & Kimmel, M. (2005). simuPOP: A forward-time population genetics simulation environment. *Bioinformatics*, *21*(18), 3686–3687. <https://doi.org/10.1093/bioinformatics/bti584>
- Plesnar-Bielak, A., Skwierzyńska, A. M., Hlebowicz, K., & Radwan, J. (2018). Relative costs and benefits of alternative reproductive phenotypes at different temperatures - Genotype-by-environment interactions in a sexually selected trait. *BMC Evolutionary Biology*, *18*(1), 1–10. <https://doi.org/10.1186/s12862-018-1226-x>
- Puig Giribets, M., García Guerreiro, M. P., Santos, M., Ayala, F. J., Tarrío, R., & Rodríguez-Trelles, F. (2019). Chromosomal inversions promote genomic islands of concerted evolution of Hsp70 genes in the *Drosophila subobscura* species subgroup. *Molecular Ecology*, *28*(6), 1316–1332. <https://doi.org/10.1111/mec.14511>
- Quilodrán, C. S., Ruegg, K., Sendell-Price, A. T., Anderson, E. C., Coulson, T., & Clegg, S. M. (2020). The multiple population genetic and demographic routes to islands of genomic divergence. *Methods in Ecology and Evolution*, *11*(1), 6–21. <https://doi.org/10.1111/2041-210X.13324>

- R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.r-project.org/>
- Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., Noor, M. A. F., Mehlig, B., & Westram, A. M. (2017). Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, 30(8), 1450–1477. <https://doi.org/10.1111/jeb.13047>
- Roesti, M., Gavrillets, S., Hendry, A. P., Salzburger, W., & Berner, D. (2014). The genomic signature of parallel adaptation from shared genetic variation. *Molecular Ecology*, 23(16), 3944–3956. <https://doi.org/10.1111/mec.12720>
- Rogers, A., & Pruu Gel-Bennett, A. (2000). Evolving Populations with Overlapping Generations. In *Theoretical Population Biology* (Vol. 57). www.idealibrary.com
- Rönkä, K., Valkonen, J. K., Nokelainen, O., Rojas, B., Gordon, S., Burdfield-Steel, E., & Mappes, J. (2020). Geographic mosaic of selection by avian predators on hindwing warning colour in a polymorphic aposematic moth. *Ecology Letters*, 23(11), 1654–1663. <https://doi.org/10.1111/ele.13597>
- Rundle, H. D., & Nosil, P. (2005). Ecological speciation. *Ecology Letters*, 8(3), 336–352. <https://doi.org/10.1111/j.1461-0248.2004.00715.x>
- Schluter, D. (2009). Evidence for ecological speciation and its alternative. *Science*, 323(5915), 737–741. <https://doi.org/10.1126/science.1160006>
- Sedghifar, A., Brandvain, Y., & Ralph, P. (2016). Beyond clines: lineages and haplotype blocks in hybrid zones. *Molecular Ecology*, 25(11), 2559–2576. <https://doi.org/10.1111/mec.13677>
- Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G. P., Bank, C., Brännström, Å., Brelsford, A., Clarkson, C. S., Eroukhmanoff, F., Feder, J. L., Fischer, M. C., Foote, A. D., Foote, A. D., Franchini, P.,

- Jiggins, C. D., ... Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics*, 15(3), 176–192. <https://doi.org/10.1038/nrg3644>
- Sendell-Price, A. T., Ruegg, K. C., Anderson, E. C., Quilodrán, C. S., Van Doren, B. M., Underwood, V. L., Coulson, T., & Clegg, S. M. (2020). The Genomic Landscape of Divergence Across the Speciation Continuum in Island-Colonising Silvereyes (*Zosterops lateralis*). *G3: Genes, Genomes, Genetics*, 10(9), 3147–3163. <https://doi.org/10.1534/g3.120.401352>
- Shafer, A. B. A., & Wolf, J. B. W. (2013). Widespread evidence for incipient ecological speciation: A meta-analysis of isolation-by-ecology. *Ecology Letters*, 16(7), 940–950. <https://doi.org/10.1111/ele.12120>
- Sinervo, B., & Lively, C. M. (1996). The rock–paper–scissors game and the evolution of alternative male strategies. *Nature*, 380(6571), 240–243.
- Smithson, A. (2001). Pollinator preference, frequency dependence, and floral evolution. In J. D. Thomson & L. Chittka (Eds.), *Cognitive Ecology of Pollination: Animal Behaviour and Floral Evolution* (pp. 237–258). Cambridge University Press. <https://doi.org/DOI:10.1017/CBO9780511542268.013>
- Sodeland, M., Jorde, P. E., Lien, S., Jentoft, S., Berg, P. R., Grove, H., Kent, M. P., Arnyasi, M., Olsen, E. M., & Knutsen, H. (2016). “Islands of Divergence” in the atlantic cod genome represent polymorphic chromosomal rearrangements. *Genome Biology and Evolution*, 8(4), 1012–1022. <https://doi.org/10.1093/gbe/evw057>
- Svensson, E. I., Gómez-Llano, M. A., Torres, A. R., & Bensch, H. M. (2018). Frequency dependence and ecological drift shape coexistence of species with similar niches. *American Naturalist*, 191(6), 691–703. <https://doi.org/10.1086/697201>
- Tepolt, C. K., & Palumbi, S. R. (2020). Rapid Adaptation to Temperature via a Potential Genomic Island of Divergence in the Invasive Green Crab, *Carcinus maenas*. *Frontiers*

- in *Ecology and Evolution*, 8(November), 1–14.
<https://doi.org/10.3389/fevo.2020.580701>
- Thompson, V. (1984). Polymorphism under apostatic and aposematic selection. *Heredity*, 53(3), 677–686. <https://doi.org/10.1038/hdy.1984.126>
- Tishkoff, S. A., Reed, F. A., Ranciaro, A., Voight, B. F., Babbitt, C. C., Silverman, J. S., Powell, K., Mortensen, H. M., Hirbo, J. B., & Osman, M. (2007). Convergent adaptation of human lactase persistence in Africa and Europe. *Nature Genetics*, 39(1), 31–40. <https://doi.org/10.1038/ng1946>
- Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, 38(6), 1358–1370. <https://doi.org/10.2307/2408641>
- Wolf, J. B. W., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, 18(2), 87–100. <https://doi.org/10.1038/nrg.2016.133>
- Woodgate, J. L., Buchanan, K. L., Bennett, A. T. D., Catchpole, C. K., Brighton, R., & Leitner, S. (2014). Environmental and genetic control of brain and song structure in the zebra finch. *Evolution*, 68(1), 230–240. <https://doi.org/10.1111/evo.12261>
- Wu, C. I., & Ting, C. T. (2004). Genes and speciation. *Nature Reviews Genetics*, 5(2), 114–122. <https://doi.org/10.1038/nrg1269>
- Xia, X. (2013). Codominance. In Stanley Maloy & Kelly Hughes (Eds.), *Brenner's Encyclopedia of Genetics* (2nd ed., pp. 63–64). Academic Press. <https://doi.org/10.1016/B978-0-12-374984-0.00278-3>
- Yeaman, S., Aeschbacher, S., & Bürger, R. (2016). The evolution of genomic islands by increased establishment probability of linked alleles. *Molecular Ecology*, 25(11), 2542–2558. <https://doi.org/10.1111/mec.13611>

2.7. Figures and Tables

Table 1 – **List of parameters**

Parameter	Definition	Value
N	Population size	400
n_L	Number of simulated loci	300
n_a	Number of additive loci	50
b_0	Maximum offspring number	6
b_1	Phenotypic optima	Pop. 1 = 0.25; Pop. 2 = 0.75 (divergent selection models only)
b_2	Variance of fitness curve	0.5
b_3	Density dependence	Pop.1 = 0.01; Pop.2 = 0.005 (divergent selection models only)
σ_{env}	Stochastic environmental component	0.01
σ_{dem}	Stochastic demographic component	1
k	Number of alleles per locus (breeding values)	20
	Sex ratio	0.5
μ	Mutation rate	0
	Linked loci position	60:69, 150:159, 230:239
ρ	Linkage	0.0001
t	Number of generations	2000
m_{ij}	Gene flow (Migration rate) † Figure S2	0, 0.005999 (constant migration rate)

0 to 0.01(intermittent migration rate)		
F ‡	Frequency dependent selection strength	2%
K ‡	Number of bins	10
G §	Genotype-environment interaction (GxE)	5,10,15,20,25

† Migration is only present for the migration specific models

‡ F and K in eq. 3-5

§ Value of G in eq.6

Table 2 – Scenarios simulated using the glads model

G-P map†	Selection	Divergent selection		Parallel selection	
Additive		Positive	frequency	Positive	frequency
		dependent selection		dependent selection	
Additive		Negative	frequency	Negative	frequency
		dependent selection		dependent selection	
Non-additive		Dominance (medium & strong)		Dominance (medium & strong)	
Additive		Genotype-environment interaction		Genotype-environment interaction	
Additive		Gene flow (constant vs intermittent)		Gene flow (constant vs intermittent)	

† Genotype-Phenotype map

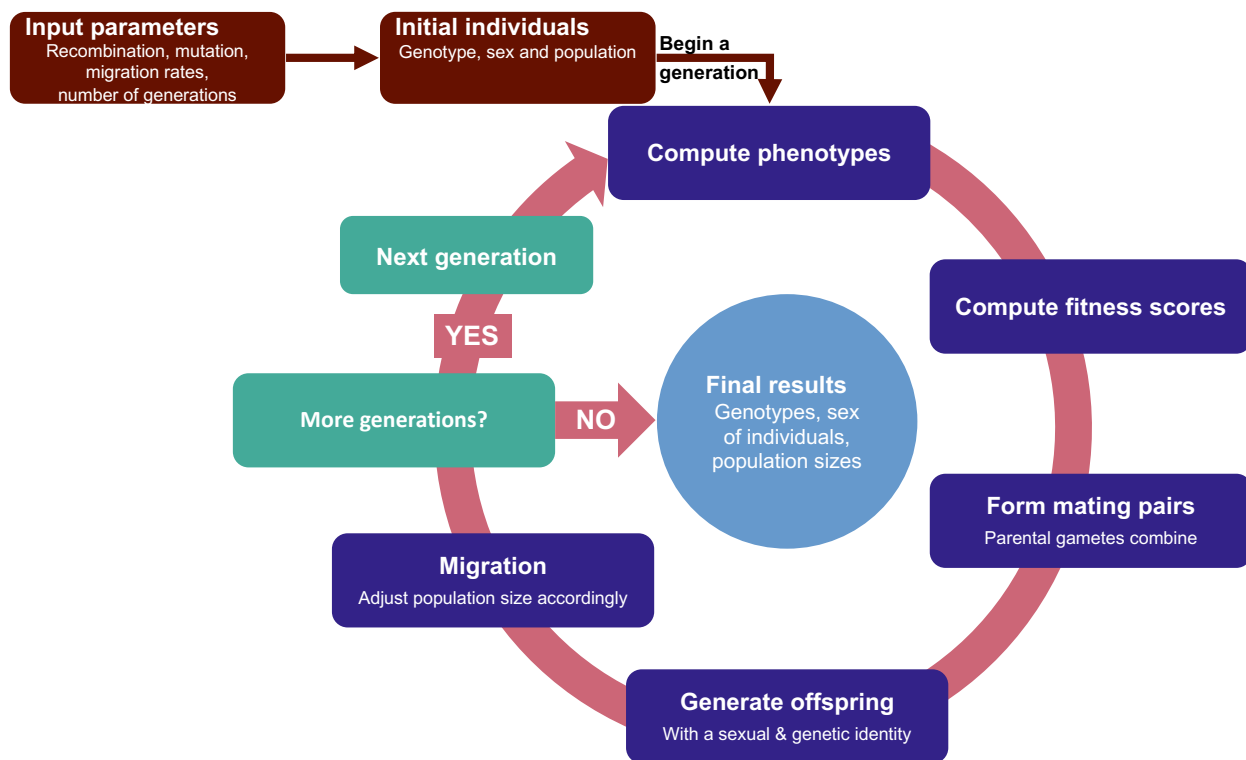


Figure 1. Summary of the ‘*glads*’ framework showing each of the main steps in the modelling approach. The boxes in brown represent initialisation of the model. The cycle represents a given time-step iteration, with the boxes indicating the different computing steps for each generation. The box “More generations?” represents a condition variable stating either the running of another generation (“Next generation” box) or the end of simulations (sky-blue circle, representing the output from which the genomic landscape is computed). Adapted from Quilodr  n et al., (2020).

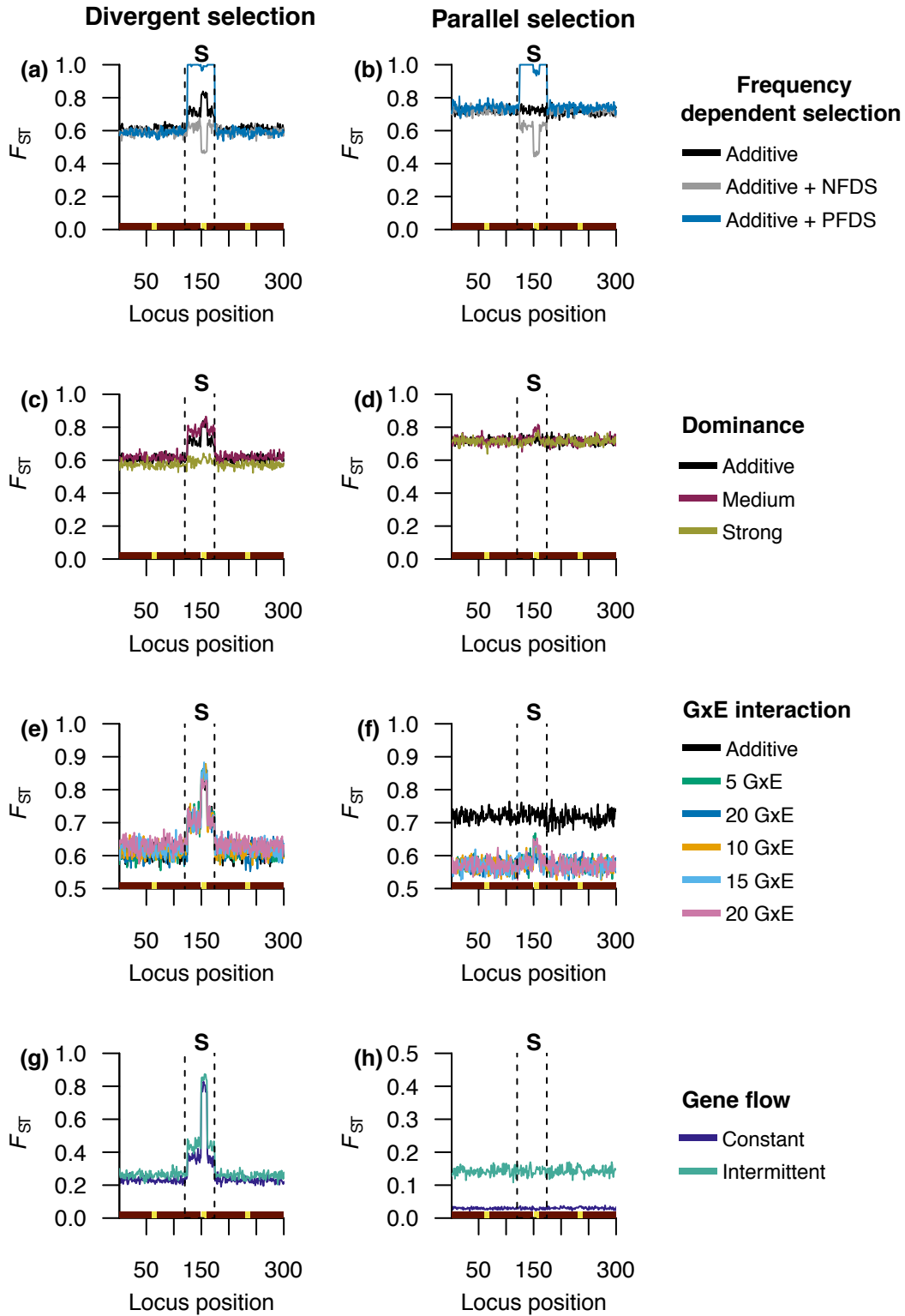


Figure 2. The genomic landscape of divergence between two populations. The output is summarised as the average F_{ST} from 100 simulations of 1000 generations each. The F_{ST}

is plotted from 0 to 1, except for e-f where the F_{ST} is plotted from 0.5 to 1. Locus position represented on the x-axis is shaded red for unlinked loci and yellow for linked loci, with loci under selection (S) that contribute to phenotypic trait values indicated by region between the vertical dashed lines. The left-hand panels show divergent selection, and the right-hand panels, parallel selection. For each of these selection regimes, graphs show the addition of: a-b) frequency-dependent selection; c-d) varying dominance strength; e-f) varying genotype-environmental interaction strength (G in eq.6); and g-h) varying modes of gene flow under an additive model. In a-f) black lines denote divergence levels under divergent or parallel selection and a purely additive genotype-phenotype map, for comparison with modelled additions.

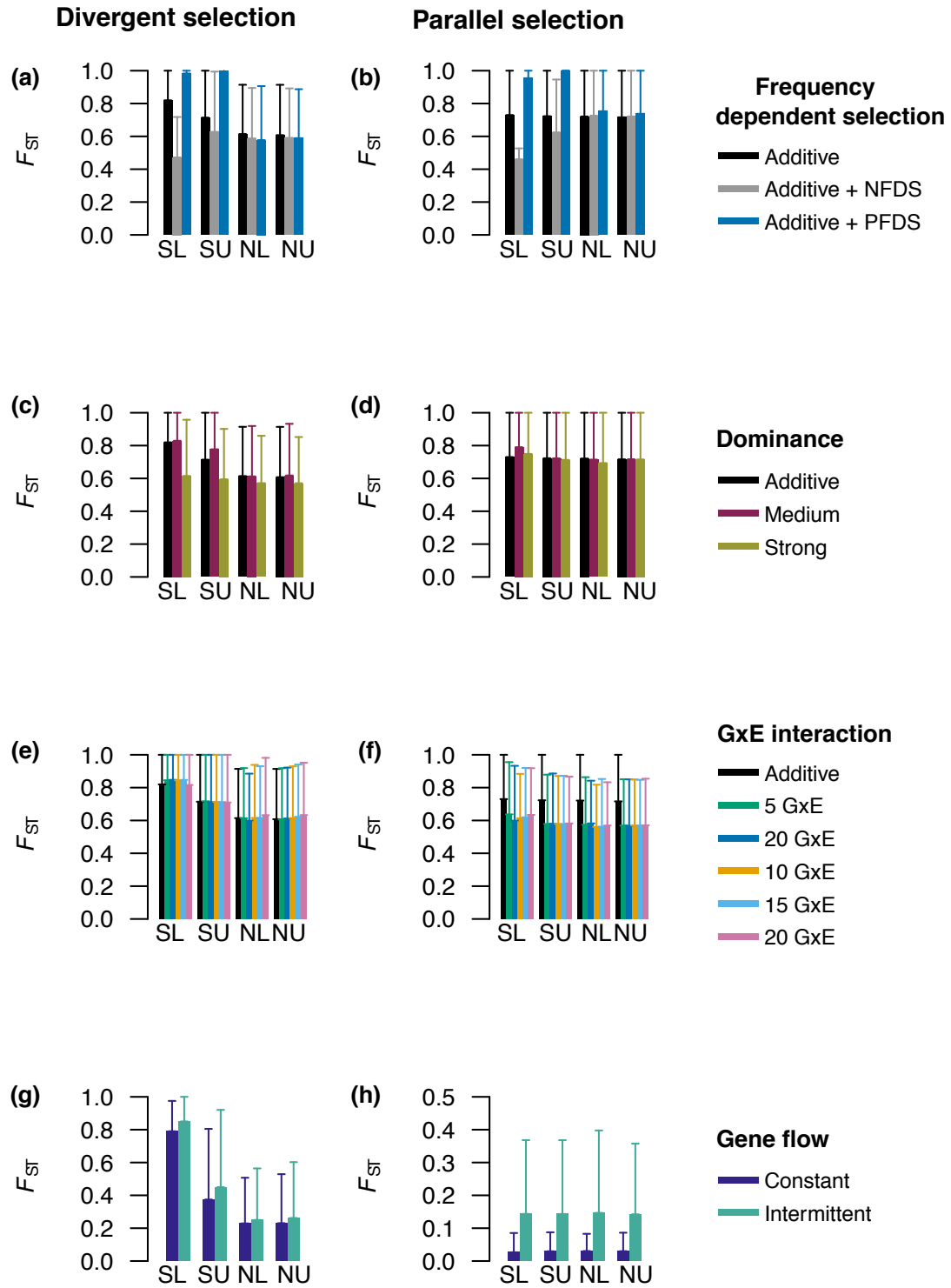


Figure 3. Mean F_{ST} for different loci types. Mean F_{ST} is given for selected linked (SL), selected unlinked (SU), neutral linked (NL) and neutral unlinked (NU) loci calculated from 100 simulations of 1000 generations, with error bars denoting the 95% quantile. The

left-hand panels show divergent selection, and the right-hand panels, parallel selection. For each of these selection regimes, graphs show the addition of: a-b) frequency-dependent selection; c-d) varying dominance strength; e-f) varying genotype-environmental interaction strength (G in eq.6); and g-h) gene flow. In a-f) the black bars denote divergence levels under divergent or parallel selection and a purely additive genotype-phenotype map, as a basis for comparison with modelled additions. The bottom panels include different types of gene flow to the same additive model.

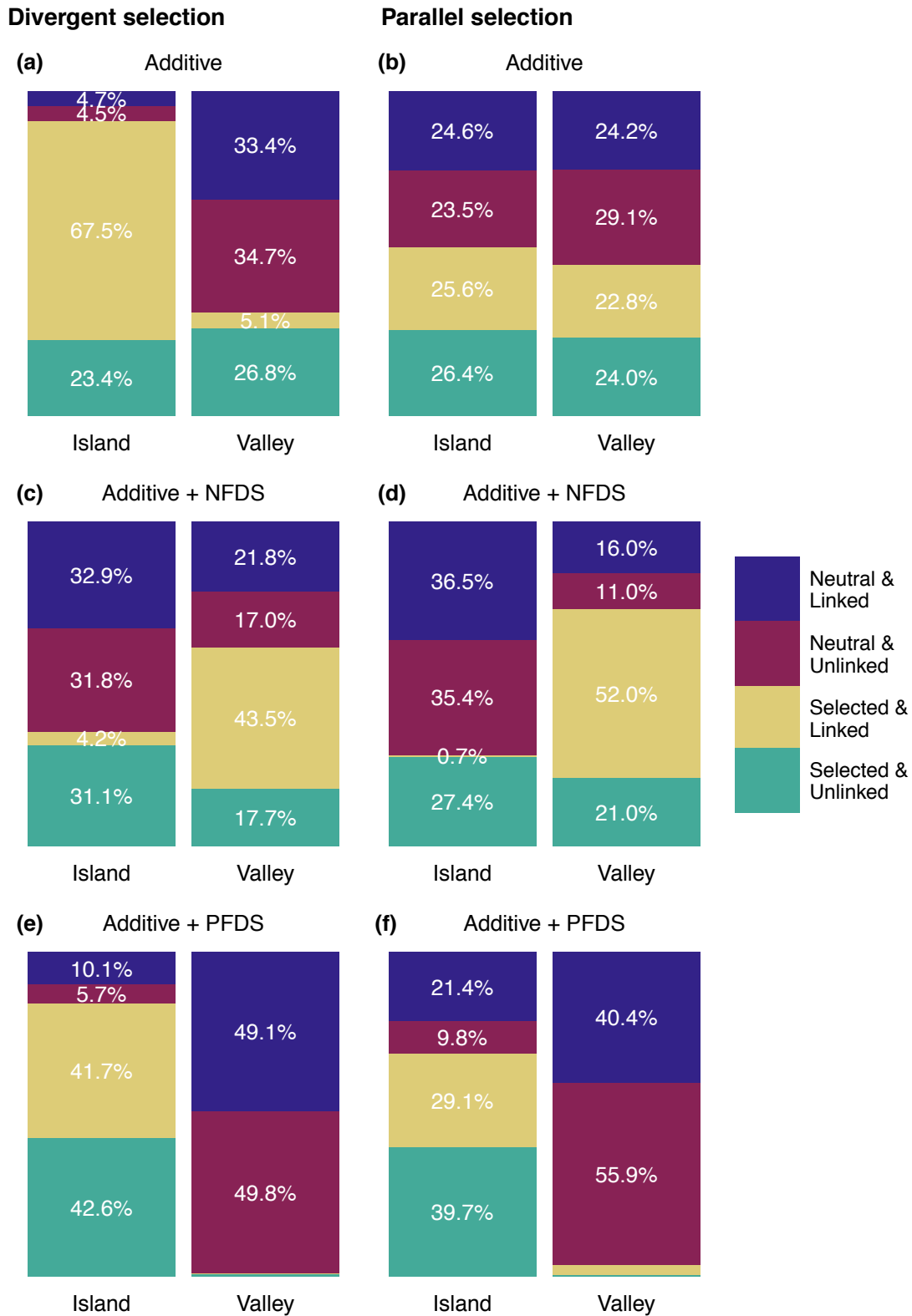


Figure 4. Proportions of islands and valleys observed on the genomic landscape of divergence under different type of selection, identified by using a hidden Markov model (100 independent simulations of 1000 generations each). The effect of divergent (left

column) and parallel selection (right column) was analysed with: (a,b) an additive genotype phenotype map; (c,d) with addition of negative frequency dependent selection (NFDS); and (e,f) with addition of positive frequency dependent selection (PFDS). The proportions of islands and valleys were computed for neutral and selected loci under linked and unlinked conditions. Under models with PFDS for selected linked and unlinked loci, the proportion of islands and valleys were 0.3% and 0.7% under divergent selection (panel e), and 3.3% and 0.4% under parallel selection, respectively (panel f).

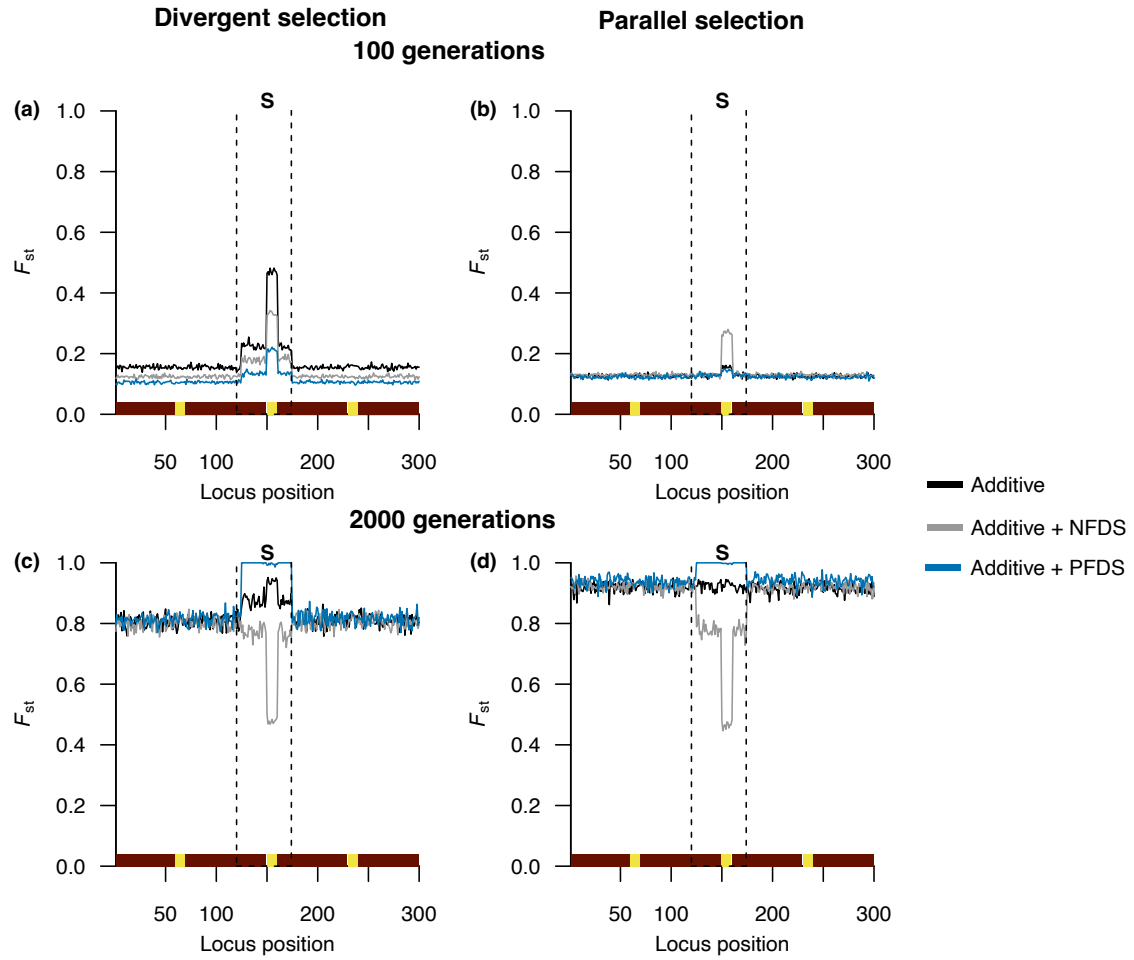


Figure 5. The genomic landscape of divergence with frequency dependent selection summarised as the average F_{ST} from 100 simulations after: a-b) 100 generations and c-d) 2000 generations. The left-hand panels show divergent selection, and the right-hand panels, parallel selection. Locus position represented on the x-axis is shaded red for unlinked loci and yellow for linked loci, with loci under selection (S) that contribute to phenotypic trait values indicated by region between the vertical dashed lines. The black lines denote divergence levels under divergent or parallel selection without frequency dependent selection, for comparison.

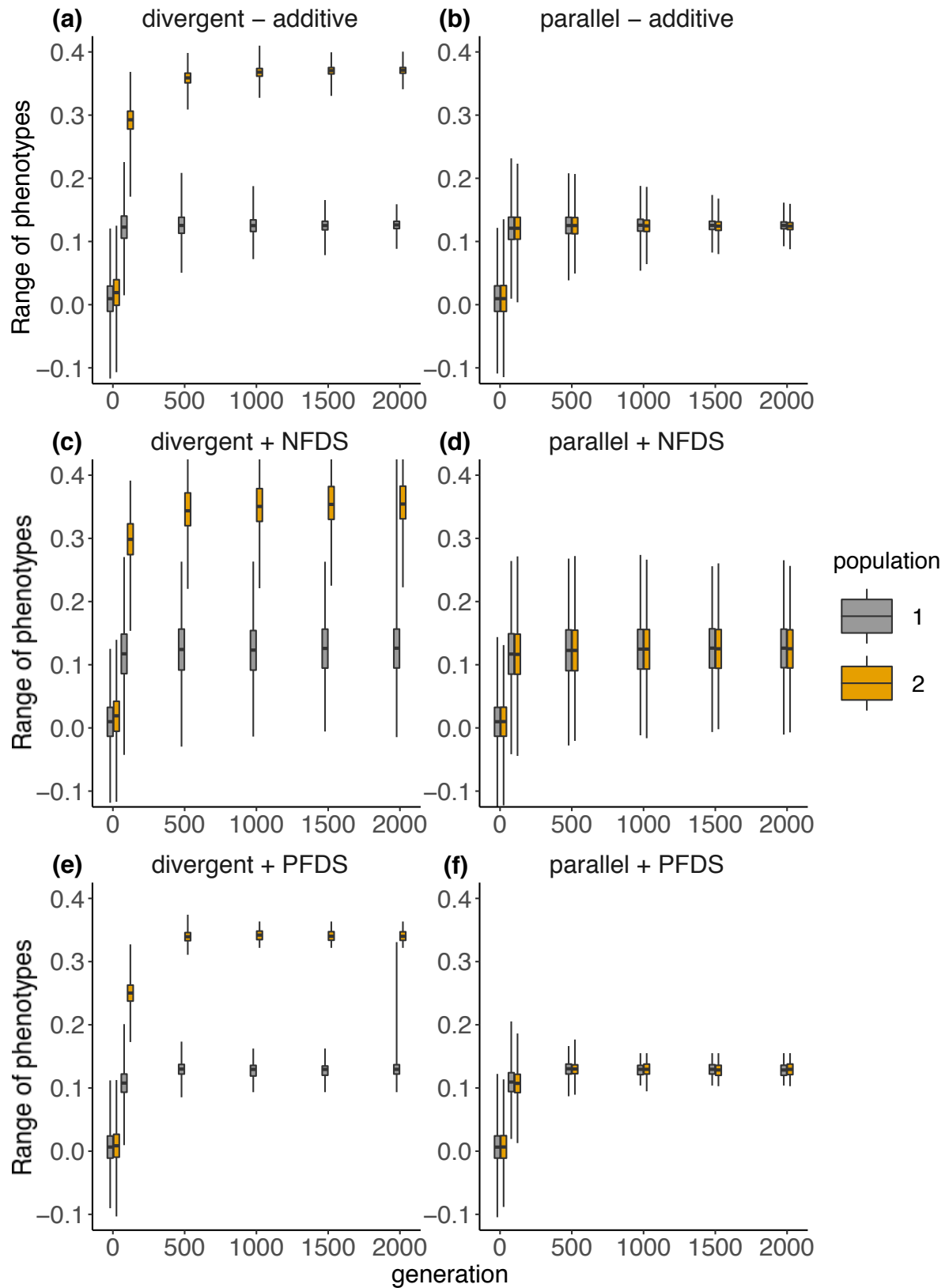


Figure 6. The phenotypic range occupied by individuals from two populations across 2000 generations. The phenotypes are standardised by dividing by the number of additive

loci (n_a). The effect of divergent (left column) and parallel selection (right column) are shown with: (a,b) an additive genotype phenotype map; (c,d) with addition of negative frequency dependent selection (NFDS); and (e,f) with addition of positive frequency dependent selection (PFDS). The whiskers represent the minimum and maximum phenotype, and the bars 25%, 50% and 75% quantiles.

2.8 Supporting Information

Supporting information for:

The effect of divergent and parallel selection on the genomic landscape of divergence

Hisham A.A. Ali, Tim Coulson, Sonya M. Clegg, Claudio S. Quilodrán

Table of Contents:

Figure S1	Page 2
Figure S2	Page 3
Figure S3	Page 4
Figure S4	Page 5
Figure S5	Page 6
Appendix S1	See .R file

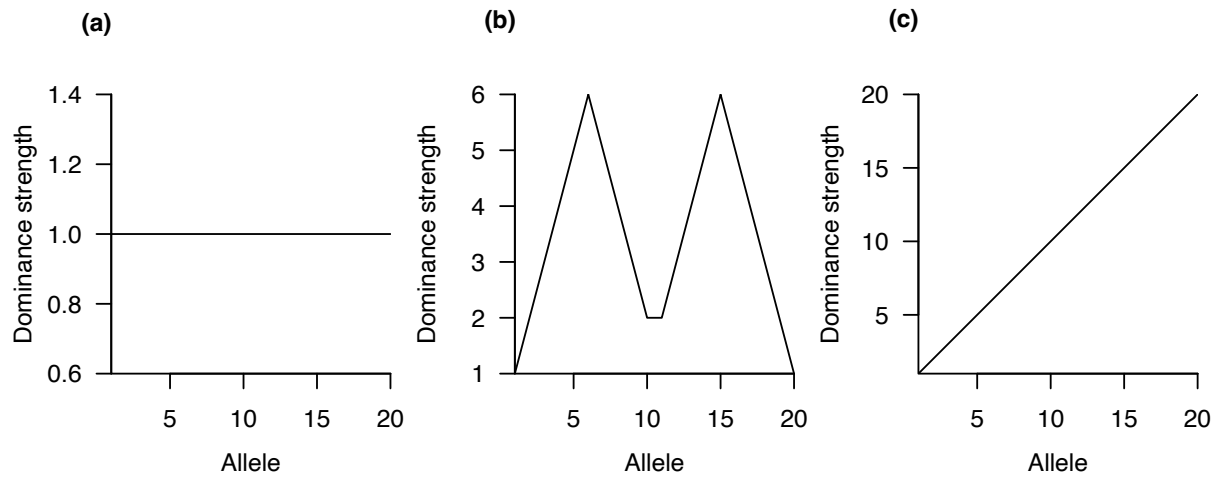


Figure S1. Three explored scenarios of dominance strength. The dominance strength was varied by using different dominance hierarchies for the 20 simulated alleles (k): (a) additive (b) medium dominance (c) strong dominance. The x-axis shows the 20 possible alleles on the loci participating in the genotype-phenotype map. The y-axis shows the hierarchy position for that allele.

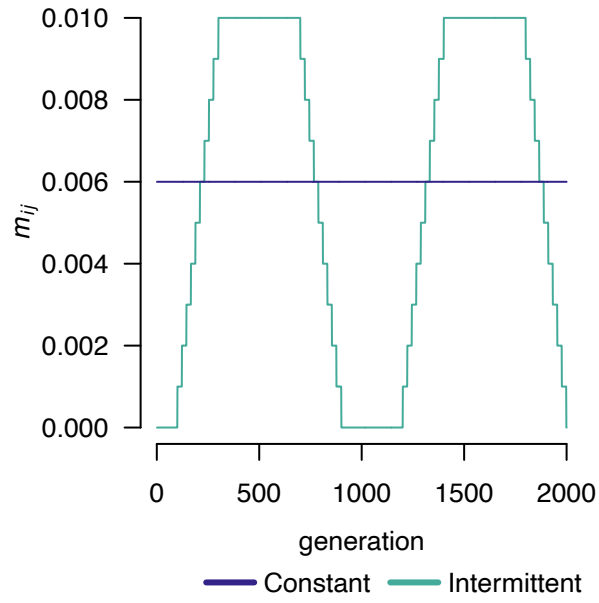


Figure S2. Two explored scenarios of gene flow. (i) constant gene flow over time (purple line), and (ii) intermittent periods with gene flow (green line). The migration rate values are chosen to obtain equivalent number of migrants at the end of the simulations.

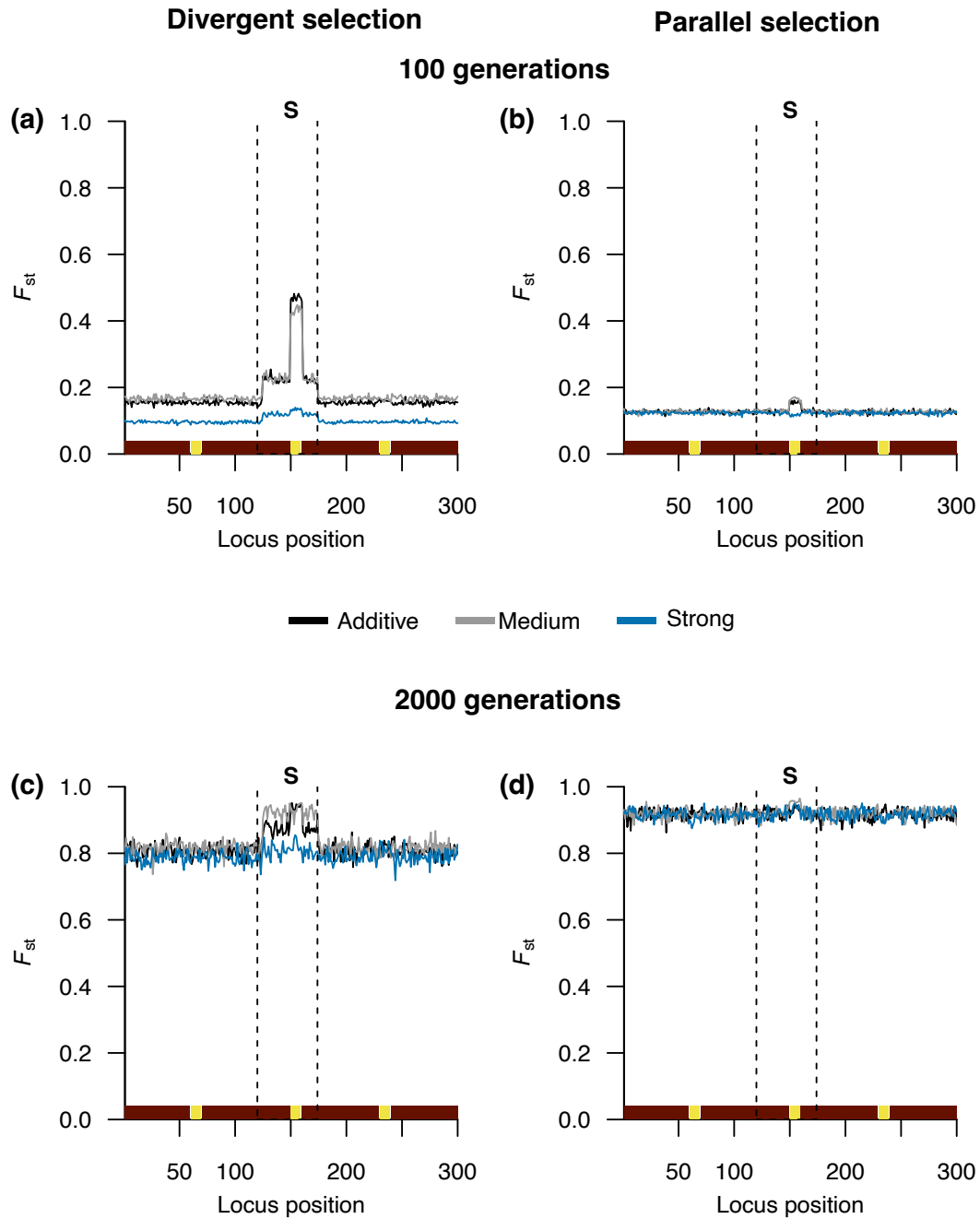


Figure S3. The genomic landscape of divergence with different dominance strengths, summarised as the average F_{ST} from 100 simulations after: a-b) 100 generations and c-d) 2000 generations. The left-hand panels show divergent selection, and the right-hand panels, parallel selection. Locus position represented on the x-axis is shaded red for unlinked loci and yellow for linked loci, with loci under selection (S) that contribute to

phenotypic trait values indicated by region between the vertical dashed lines. The black lines denote divergence levels under divergent or parallel selection without dominance, as a basis for comparison with medium and strong dominance (grey and blues lines respectively).

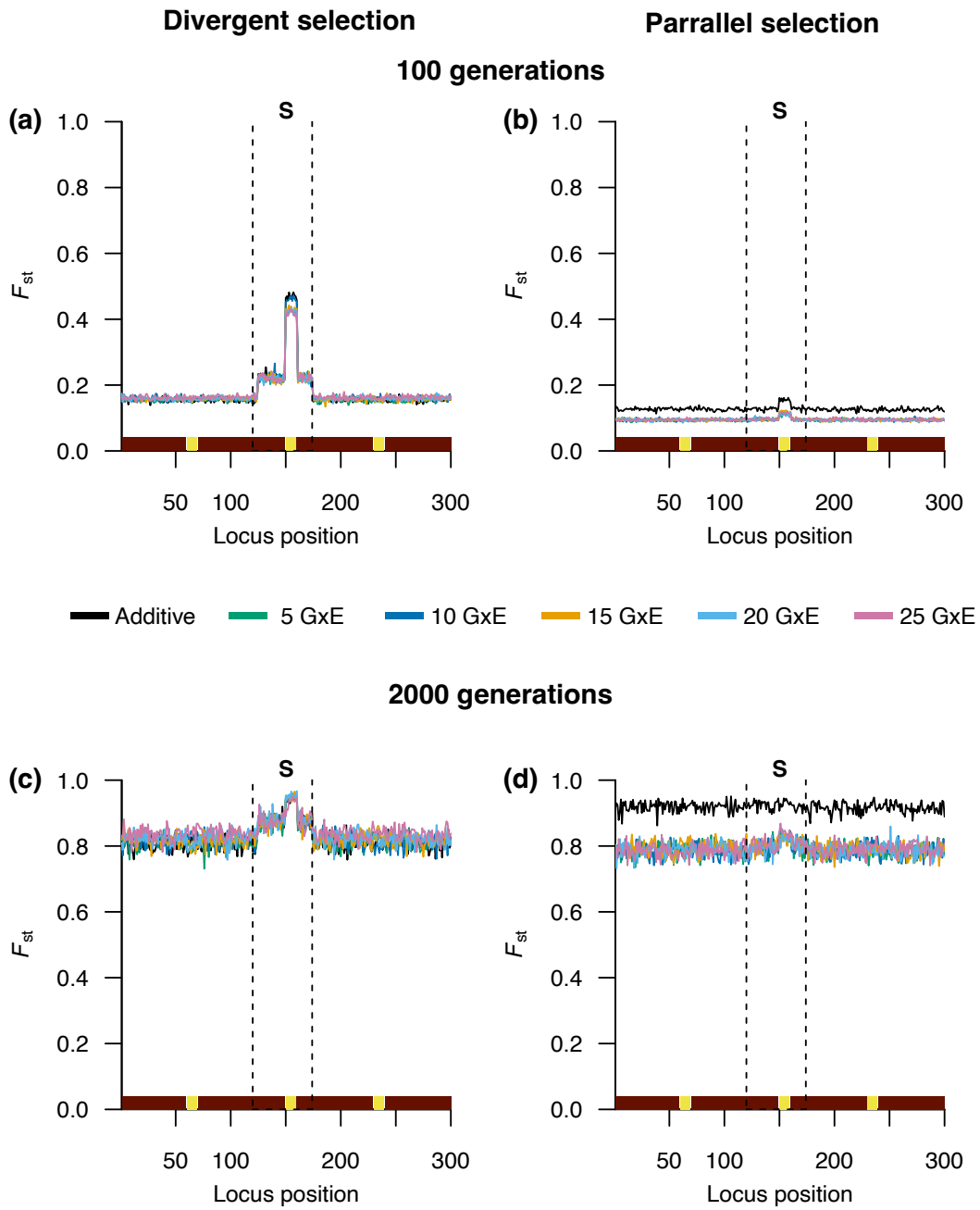


Figure S4. The genomic landscape of divergence with different genotype-environmental (GxE) interaction strengths, summarised as the average F_{ST} from 100 simulations after: a-b) 100 generations and c-d) 2000 generations. The left-hand panels show divergent selection, and the right-hand panels, parallel selection. Locus position represented on the x-axis is shaded red for unlinked loci and yellow for linked loci, with loci under selection (S) that contribute to phenotypic trait values indicated by region between the vertical dashed lines. The black lines denote divergence levels under divergent or parallel selection without GxE, as a basis for comparison with the coloured lines represent different GxE strength ranging from 5 to 25 G (see Methods).

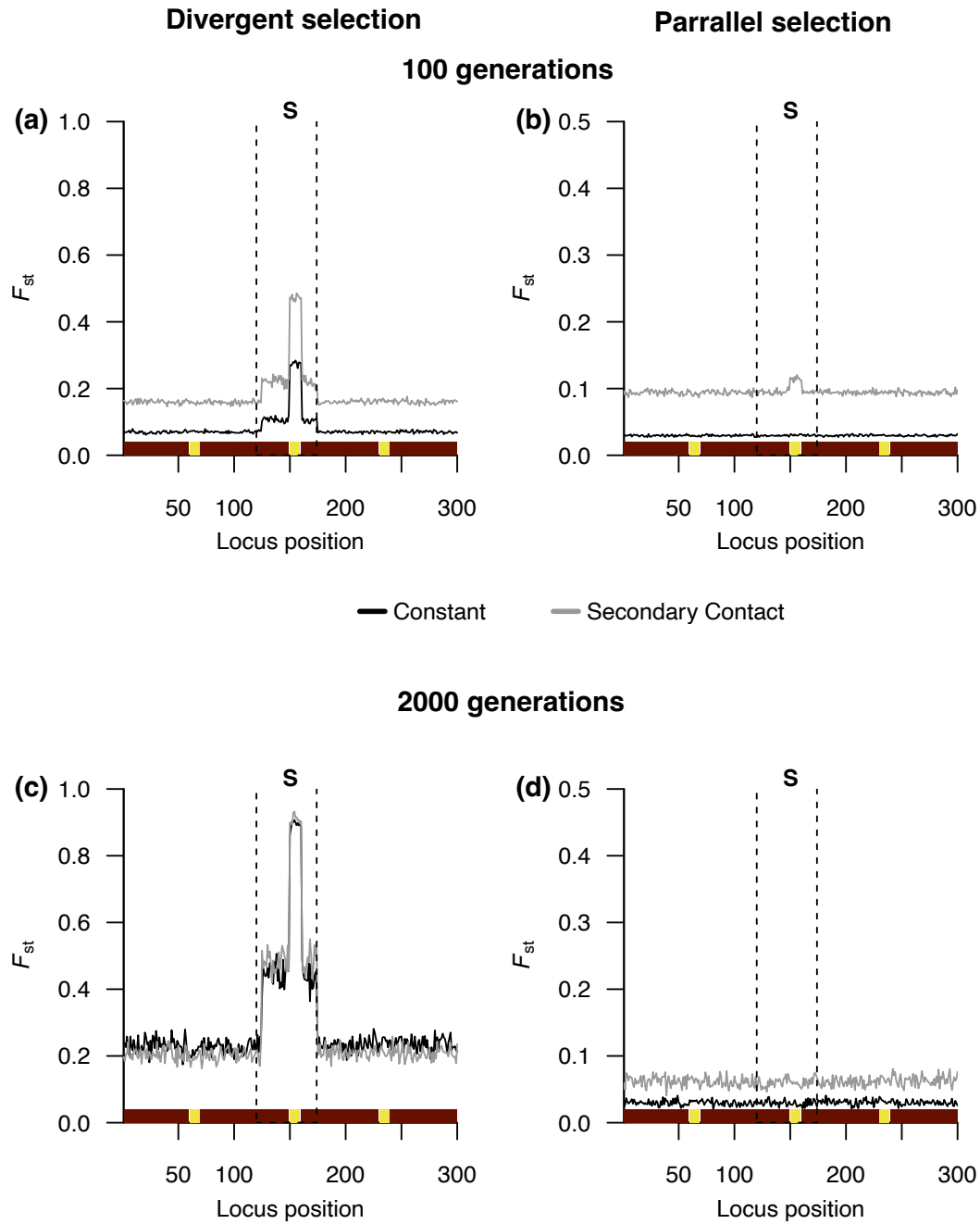


Figure S5. The genomic landscape of divergence with different types of gene flow, summarised as the average F_{ST} from 100 simulations after: a-b) 100 generations and c-d) 2000 generations. The left-hand panels show divergent selection, and the right-hand panels, parallel selection. Locus position represented on the x-axis is shaded red for unlinked loci and yellow for linked loci, with loci under selection (S) that contribute to

phenotypic trait values indicated by region between the vertical dashed lines. The black lines represent a constant level of gene flow and the grey lines represent intermittent gene flow (see Methods).

3. The effect of Dobzhansky-Muller incompatibilities on the genomic landscape of divergence under parallel and divergent selection

Abstract

Genomes diverge in a heterogeneous fashion, generating the genomic landscape of divergence. However, understanding how different evolutionary processes shape the genomic landscape of divergence is challenging. Dobzhansky-Muller incompatibilities (DMIs) describe the negative epistatic interactions that can arise in an individual when incompatible parental mutations are combined. There is a body of theoretical and empirical work that shows these interactions can reinforce population divergence through selection against hybrids, which may ultimately result in speciation. They may be particularly effective at causing population divergence when similar (parallel) ecological selection occurs but may also play a role when divergent selection is operating. However, under gene flow, the formation of DMIs at neutral loci and loci under parallel selection is challenging. How Dobzhansky-Muller incompatibilities (DMIs) contribute to the genomic landscape concurrently with other evolutionary processes has not been explored. Using an individual-based framework that includes complex demographic and genetic interactions, we modelled how DMIs influence the genomic landscape of divergence. We considered two diverging populations defined by two quantitative phenotypic traits under three combinations of selection regimes: i) both traits evolving under parallel selection; ii) both traits evolving under divergent selection; and iii) one trait evolving under parallel selection and the other under divergent selection. We showed that DMIs allow genetic

divergence to accumulate despite medium levels of migration between diverging populations, under simulations of parallel and divergent selection. We also showed that the effect of selection on the genomic landscape is counterbalanced when both parallel and divergent selection concurrently affect epistatic interactions. Our results highlight the influence of DMIs on accumulating divergence with migration, as well as the complex patterns that could arise on the genomic landscape of divergence under the concurrent influence of different types of selection.

Keywords: divergence, Dobzhansky-Muller incompatibility, epistasis, genomic landscape, mutation, natural selection

3.1. Introduction

The genomic landscape of divergence describes the heterogenous pattern of differentiation across the genome when comparing diverging populations (Gavrilets, 2014; Michel et al., 2010). Initial interpretations of the processes underlying observed patterns, such as the joint role of selection and drift in producing genomic islands of divergence, have proved to be oversimplifications (Campbell et al., 2018; Quilodr  n, Ruegg, et al., 2020a; Ravinet et al., 2017). Yet empirical and modelling studies have narrowed down the combinations of circumstances and processes that can produce particular patterns (Andrello & Manel, 2015; Quilodr  n, Ruegg, et al., 2020a; Wolf & Ellegren, 2017). However, the effects of multiple processes on the genomic landscape of divergence as speciation progresses along a continuum remain to be examined, and one of these is the influence of Dobzhansky-Muller incompatibilities (DMIs).

DMIs describe the formation of negative epistatic interactions between alleles that arise on different genomic backgrounds, that when combined result in decreased hybrid fitness: a decreased fitness for an individual that inherits DMI mutations from different lineages (Corbett-Detig et al., 2013). As a consequence, divergence and potentially post-zygotic reproductive isolation, can be generated by DMIs (Chou et al., 2022; Matute et al., 2010). DMIs can form under both parallel and divergent selection regimes (Schluter & Conte, 2009). Under parallel selection, populations face similar ecological conditions, and similar phenotypes are favoured. Different, but incompatible, mutations that achieve a similar phenotype can form in each population, resulting in the accumulation of DMIs; thus, reducing gene flow (Chou et al., 2022), and potentially resulting in mutation order speciation (Schluter, 2009). However, under gene flow this is challenging, as divergence is hindered by the movement of beneficial mutations between populations (Schluter, 2009). Under divergent selection, DMIs can accumulate as ecologically divergent phenotypes are favoured in different environments (Rundle & Nosil, 2005), regardless of whether gene flow is occurring or not (Schluter, 2009).

Modelling approaches can be used to provide insights into how DMIs influence the genomic landscape of divergence under parallel and divergent selection. Several studies have used Fisher's geometric model to explore the effects of varying epistatic interactions and the environmental optima on reproductive isolation caused by DMI, in terms of F1 hybrid fitness (Turelli and Orr, 2000; Fraïsse et al., 2016; Simon, Bierne and Welch, 2018; Yamaguchi and Otto, 2020). Different selection strengths have also been simulated using Fisher's geometric model by use of a multi-variate normal distribution (Waxman and Welch, 2005; Martin and Lenormand, 2006; Chevin, Decorzent and Lenormand, 2014; Fraïsse et al., 2016). Fisher's geometric model derives probabilities for the fitness effect of mutations under different mutational effect sizes, the number of

traits, and the distance to the fitness optima (Fisher, 1930). Additionally, Fisher's geometric model inherently implements epistasis and so can facilitate DMIs (Fraïsse et al., 2016). These studies provide insights into the effect of DMIs on hybrid fitness, but do not reveal how the genomic landscape of divergence changes during the speciation process. Furthermore, the effect of parallel versus divergent selection on DMIs has not been explored.

The genomic landscape of divergence has been modelled in a number of ways. Early methods used either a single or a few loci with selection defined as a single parameter (Charlesworth, Nordborg and Charlesworth, 1997; Feder and Nosil, 2009, 2010; Feder et al., 2012; Sedghifar, Brandvain and Ralph, 2016). Larger stretches of genetic data, up to chromosome-level, have been modelled with coalescence-based approaches, which are less informative for small non-panmictic populations under selection compared to large populations without selection (Currat et al., 2015). An alternative to the coalescent approach is forward-in-time individual-based models (IBMs) that enable simulation of complex demographic and genetic dynamics that have been applied to small populations with few loci (Hoban et al., 2012). However, the computational requirements needed for simulating large chromosomal level genetic data and large populations using IBMs are now tractable (Currat et al., 2015).

We use the 'glads' individual-based framework (Quilodr  n, Ruegg, et al., 2020a) to model how DMIs shape the genomic landscape of divergence under different selection regimes. The glads model is a highly flexible approach that simulates the dynamic formation of the genomic landscape of divergence, under various evolutionary and demographic parameters, for genetically defined populations. We also make use of the approaches developed for Fisher's geometric model to simulate DMIs (Fisher, 1930). First, we extend the 'glads' fitness function to a modified bivariate normal distribution,

which allows the simulation of different strengths and types of selection (i.e. parallel or divergent) on two traits within a simulated individual, instead of a single trait. Second, we modify the mutation rate parameter to incorporate DMIs into the model, such that mutations that form are either DMI or non-DMI mutations. Third, we develop an epistatic map and count-based approach in conjunction with a parameterizable function that calculates the fitness depression as a consequence of DMIs in hybrids. With these modifications, we can utilise the *glads* model to examine their effects on the genomic landscape of divergence.

The framework for the model is based on an island-colonising silvereye (*Zosterops lateralis*) birds that show repeated body size increases (Clegg, Degnan, Moritz, et al., 2002; Clegg et al., 2008), but divergent bill shape evolution (Clegg, Degnan, Kikkawa, et al., 2002; Clegg & Phillimore, 2010), consistent with traits under parallel selection and divergent selection respectively. We parameterise the model using demographic and genetic information from one well-studied island silvereye population: Heron Island, Australia. We apply a fitness function to simulate populations with two traits under selection, applying either parallel or divergent selection to both traits, or applying parallel selection to one trait and divergent selection to the other, and including the formation of DMIs to examine their impact on the genomic landscape of divergence.

3.2. Methods

3.2.1 Description of the model

Simulations were performed using a modified version of the Genomic Landscape of Divergence Simulations (*'glads'*) framework developed by Quilodr  n et al. (2020). This

is an IBM that simulates the divergence of two or more populations. The purpose of this approach is to provide insights about the concurrent effect of genetic and demographic processes on the development of patterns across the genomes of divergent populations.

The individuals within populations have three hierarchical levels: the genotype, phenotype, and fitness. The genotype is the diploid allelic composition of an individual across a predefined number of loci, some of which affect the phenotype. Each allele at a locus that contributes to the phenotype is assigned a breeding value. The individual phenotype (z) (equation 1) is calculated as the sum of the breeding values (b_v) for all selected loci (n_a) along with a stochastic environmental component (ε_{env}). The environmental component is defined at the population level with a population mean (0) and standard deviation (σ_{env}).

$$z = \sum_{v=1}^{n_a} b_v + \varepsilon_{env}(0, \sigma_{env}) \quad (1)$$

The fitness (ω) of an individual is calculated by applying the phenotype to a modified Gaussian function (equation 2). This function is defined for each population by the maximum number of generated offspring (b_0), the phenotypic optima with the maximum fitness (b_1), and the variance of the phenotypes (b_2). Natural selection acting between populations can be implemented either under parallel selection (where populations have equal phenotypic optima) or divergent selection (where populations have different phenotypic optima).

The fitness function also includes a density-dependent component (b_3) to prevent exponential growth of the population size (N). A stochastic demographic component (ε_{dem}) with a predefined variance (σ_{dem}) and mean (0) is also included. The model

progresses in discrete steps for each generation, where individuals reproduce sexually at a rate according to their fitness.

$$\omega = b_0 e^{-\frac{1}{2}(\frac{4z-b_1n_a}{b_2n_a})^2} - b_3N + \varepsilon_{dem}(0, \sigma_{dem}) \quad (2)$$

The populations are made of genetically defined individuals with a sexual identity, mutation and recombination rate. The number of mutations per site per generation for bi-allelic loci are generated using a binomial distribution where the probability of a mutation is given by the mutation parameter (μ). The crossing over points are drawn using a Poisson process, where the probability of a recombination event is given by the recombination rate parameter (ρ). The populations are also assigned demographic parameters: a between-population migration rate (m_{ij}) of sexually mature individuals, and within-population sex ratios. Further details of the ‘*glads*’ model are outlined in Quilodr  n et al. (2020).

3.2.2 Simulating Dobzhansky-Muller Incompatibilities (DMIs)

New mutations that generate reproductive incompatibilities between populations (i.e. DMIs) were introduced using the additive fitness function (equation 2). We develop two ways in which DMIs can arise at both neutral and selected loci in the simulation: i) by a fixed rate per site or ii) by using an epistatic map. The first method can be used to simulate the DMI rate of a particular system without knowing the location of the DMIs. This method calculates the number of mutations that are DMIs per site per generation using a binomial distribution, where the probability a mutation will form a DMI in a hybrid individual is given by p_{DMI} (henceforth referred to as DMI mutations). As a consequence

of gene flow, individuals can inherit DMI mutations originating from different populations. DMI mutations within their population of origin do not decrease the fitness, but through negative epistatic interactions with DMI mutations from other populations, they decrease the fitness of the individual. In these individuals, the number of DMI mutations from each parental population (λ_i and λ_j) are calculated and applied to a user defined fitness function $h(\lambda_i, \lambda_j)$, giving the realised fitness (equation 3).

$$\omega = b_0 e^{-\frac{1}{2} \left(\frac{4z - b_1 n_a}{b_2 n_a} \right)^2} - b_3 N + \varepsilon_{dem} + h(\lambda_i, \lambda_j) \quad (3)$$

The second approach facilitates the simulation of precise epistatic interactions between particular loci, which may be known from empirical data. This approach implements a user defined epistatic map describing which loci generate negative epistatic incompatibilities. The epistatic map is defined by a two-dimensional array, with each dimension representing a population, and each row and column representing a locus. Note, this approach can be extended to more than two populations, with the addition of a dimension to the array for each additional population. Like the previous approach, negative epistatic interactions only form between DMI mutations with different population origins. Therefore, by comparing DMI mutation positions to this array, the number of negative epistatic interactions for each individual (λ_i) is calculated and applied to the user defined fitness function $h(\lambda_i)$, giving the realised fitness (equation 4). In this study, we are primarily interested in the general effects of DMIs on the genomic landscape, rather than for a specific epistatic map, but we provide an example of the epistatic map approach in the supplementary material (see Appendix S1).

$$\omega = b_0 e^{-\frac{1}{2} \left(\frac{4z - b_1 n_a}{b_2 n_a} \right)^2} - b_3 N + \varepsilon_{dem} + h(\lambda_t) \quad (4)$$

For the custom function $h()$ in equations 3 and 4, a negative function has to be defined to simulate negative epistatic interactions. If the custom term generates a negative realised fitness ($\omega < 0$), then the realised fitness of the individual is simulated as 0. A positive function can also be provided to simulate the effects of positive epistatic interactions between mutations.

By using a negative custom function for $h()$, as described in the approach of Turelli & Orr (2000), we were able to model the intermediate effects of mildly deleterious DMIs, in contrast to the extremes of positive epistatic interactions where the hybrids were fitter than parental genotypes, or negative epistatic interactions that severely decreased the fitness (Simon et al., 2018). By simulating recombination, DMI mutations can independently assort between generations, in contrast to the fixed DMI transmission rate between parents and offspring of 50% (Chevin et al., 2014; Yamaguchi & Otto, 2020), thus simulating the stochastic aspect of the inheritance of DMI mutations between generations.

3.2.3 Multi-trait model

In order to model two traits within each individual, we adapted the fitness function (equation 3 & 4) to a bivariate normal distribution. The fitness for each individual ($\omega_{\alpha,\beta}$) was calculated using the additive phenotype of each trait, α and β in equation 5. The trait phenotypes were calculated by applying equation 1 to only the loci associated with that

trait; whereby n_α and n_β denoted the number of loci associated with each trait. Each population was also separately assigned parameters for each trait: first, by assigning the phenotypic optima ($b_{1\alpha}$, $b_{1\beta}$) and variance ($b_{2\alpha}$, $b_{2\beta}$) for each trait, a normal distribution is described for each population; second, by specifying the correlation between the two traits by $P \in [-1, 1]$. The effect of DMIs (by using the custom function $h()$ described in either equation 3 or 4) was applied as a 4th term to the fitness function (equation 5).

$$\omega_{\alpha,\beta} = b_0 e^{-\frac{1}{2(1-P^2)} \left[\left(\frac{\alpha - b_{1\alpha} n_\alpha}{b_{2\alpha} n_\alpha} \right)^2 - 2P \left(\frac{\alpha - b_{1\alpha} n_\alpha}{b_{2\alpha} n_\alpha} \right) \left(\frac{\beta - b_{1\beta} n_\beta}{b_{2\beta} n_\beta} \right) + \left(\frac{\beta - b_{1\beta} n_\beta}{b_{2\beta} n_\beta} \right)^2 \right]} - b_3 N + \varepsilon_{dem} + h()$$

(5)

3.2.4 Initialisation of the model

The model parameterisation was based on estimations from a silvereeye (*Zosterops lateralis*) population on Heron Island, Australia. Each population was set to simulate 400 individuals with a 50:50 sex ratio for each simulated population, mirroring the breeding population size and sex ratio of Heron Island silvereyes (Kikkawa & Wilson, 1983). All simulations were run for 1000 generations, based on a Heron Island silvereeye population age of three to four thousand years (length of time the island has been vegetated) and a generation time of three years (Clegg et al., 2008; Sendell-Price et al., 2020). Using the multi-trait fitness function (equation 5), 50 different additive loci were simulated for each of the two uncorrelated traits ($P = 0$). The remaining 200 loci did not contribute to the additive phenotype, simulating the effects of drift and recombination under neutral evolution. The allele frequency at each biallelic locus in the simulated populations were parameterised using restriction site-associated DNA sequencing (RAD-seq) data for the

approximately 11.3 Mbp chromosome 17 (Anderson et al., 2019). The sequence data was adapted for use in the ‘*glads*’ model by using the *geno2glads* function (see Appendix S1). This function generates a population of genetically defined individuals, where the frequency of heterozygotes at each locus in the simulated population is equal to the frequency of heterozygotes at each corresponding locus in the empirical sequence data.

For the simulations, we used a migration rate of either 0.0001 or 0.01. These rates were based on empirical data for the silveryeye population on Heron Island. The higher rate (henceforth referred to as the medium migration rate) was based on the empirical upper estimate of the migration rate of silveryeyes between Heron Island and neighbouring islands that harbour the same subspecies (Brook and Kikkawa, 1998; Sendell-Price et al., 2020). To approximate the lower rate of migration between Heron island and the mainland, this upper estimate was reduced by two orders of magnitude (henceforth referred to as the low migration rate) (Sendell-Price et al., 2020). Recombination points were calculated using the ‘*glads*’ *average* method at a rate of 6 cM/Mb. This rate was based on the recombination rate of the collared flycatcher (*Ficedula albicollis*) (Kawakami et al., 2014). We simulated DMI mutations using a probability approach ($p_{DMI} = 0.8$) under medium and low migration rates, to explore the effects of DMIs on the landscape of divergence under different levels of migration. Under the medium migration model, we also simulated a lower DMI probability $p_{DMI} = 0.2$, to explore the effects of DMI probability on the landscape of divergence. To accommodate the formation of 3 DMI mutations across 10% of the population per generation, when a DMI rate of $p_{DMI} = 0.8$ was simulated, an adjusted mutation rate of 6.25×10^{-4} mutations per locus per generation was used. We also simulated a threshold approach, where if the number of DMIs exceeded or was equal to the threshold, the individual was sterile (Turelli & Orr, 2000; Yamaguchi & Iwasa, 2017). For the DMI probability approach, we

used the custom negative function $h(\lambda_i, \lambda_j) = -2\min(\lambda_i, \lambda_j)$, so that for our threshold of 3 DMIs, $h(3)$ was equal to b_0 . In that case, under equation 5, the fitness of an individual with 3 DMI mutations would be 0 ($\omega_{\alpha\beta} = \in [0, 6] - 6$), and the individual would be sterile, and so would not contribute genetically to the following generation (representing post-zygotic reproductive isolation).

A conservative estimate of six was used for the maximum fitness (b_0): this was calculated by multiplying the 3-year generation time of the majority of the population with the average number of fledglings per clutch (1.9), which was then rounded to the nearest whole integer (non-breeding individuals were not considered) (Clegg et al., 2008; Kikkawa & Wilson, 1983). A density dependence strength (b_3) of 0.006 was used to maintain the population size at 400 individuals (Kikkawa & Wilson, 1983).

3.2.5 Simulations

We simulated the effects of different selection regimes on the genomic landscape of divergence. Three selection regimes were simulated using the multi-trait model: i) divergent selection for both traits, ii) parallel selection for both traits, and iii) parallel selection for trait 1 and divergent selection for trait 2. Furthermore, each model was run with and without DMIs. Each simulation was repeated 100 times, and in each replicate the starting genetic identity of individuals varied. While the allele frequencies were the same between the replicates, changing the “seed.number” in each *geno2glads* replicate varied which individuals were heterozygous for different loci in the population. We performed paired t-tests, for simulations with ($pDMI = 0.8$) and without DMIs to compare the population sizes for the 100 independent simulations of divergent and parallel selection.

The fixation index (F_{ST}) (Weir and Cockerham, 1984) was measured every 100 generations for each single nucleotide polymorphism (SNP) to quantify the genomic landscape of divergence. The simulations were performed at the University of Geneva on the “Yggdrasil” HPC cluster. F_{ST} calculations and ‘*glads*’ simulations, including population matrix generation with *geno2glads*, were carried out on the cluster using R software v3.61 (R Core Team, 2019). F_{ST} was calculated using the R package ‘*pegas*’ (Paradis, 2010). Whether a mutation was a DMI or non-DMI mutation (i.e. whether it generated epistatic interaction) did not influence F_{ST} calculations.

A Hidden Markov Model (HMM) was applied to the F_{ST} output of the simulated models to observe the effects of different DMI mutation rates and selection regimes. Equal numbers of three different types of loci (neutral, trait 1, and trait 2 coding loci) were selected to standardize the HMM. The HMM was used to distinguish three hidden states: i) background levels of differentiation; ii) genomic islands; and iii) genomic valleys. This HMM analysis was computed following the approach of Marques et al. (2016) using the ‘*HiddenMarkov*’ R package (Harte, 2017). The parameter with the highest likelihood was calculated from 1000 random initial parameter values using the Baum-Welch algorithm. Subsequently, the Viterbi algorithm was applied to these parameters to find the highest probability sequence of states across the simulated loci. Significance for these calculated states was calculated by 10,000 random permutations of genomic islands and valleys across the simulated loci (Quilodrán, Ruegg, et al., 2020a).

3.3. Results

3.3.1 Effects of DMIs on the genomic landscape of divergence

The average F_{ST} of 100 independent simulations showed a positive relationship between the probability of a DMI mutation arising (pDMI) and divergence level after 1000 generations (Fig. 1,2). Across the three selection regimes (i.e. divergent selection for both traits; parallel selection for both traits; and parallel selection for trait 1 and divergent selection for trait 2) the divergence level was similar for neutral and selected loci when DMI mutations were set to zero (Fig. 1, 2). When DMI mutations were implemented, F_{ST} was slightly higher for selected loci than neutral loci (Fig. 1,2). For simulations that included DMI mutations (pDMI=0.8), the population size was significantly larger for simulations conducted with divergent selection on the two traits ($M=58.95$, $SD= 5.78$) than simulations conducted with parallel selection on the two traits ($M= 44.27$, $SD= 5.85$); paired t-test: $t(26)= 10.57$, $p= 6.59 \times 10^{-11}$.

The Hidden Markov Model analysis showed that both genomic islands and valleys formed across the 100 independent simulations, irrespective of the selection regime and locus type (neutral loci and selected loci for traits 1 and 2) (Fig. 3). A higher proportion of genomic islands formed at selected loci (trait 1 and 2) with DMIs than at selected loci without DMIs, a difference that increased as pDMI was raised (Fig. 3). When including DMIs, the three selection regimes showed a similar percentage of genomic islands for both traits, and a higher percentage of islands compared to neutral loci, particularly when DMI mutations were highly likely to form (pDMI=0.08) (Fig. 3d-i). When DMI mutations were set to zero, a similar number of genomic islands emerged for both traits and neutral loci (Fig. 3a-c). The percentage of genomic valleys were similar for neutral

and selected loci when DMI mutations were set to zero, and where they were allocated a lower probability of forming (Fig. 3a-f). However, when the probability of DMI mutations forming was set to a high value, the genomic valleys formed more frequently at neutral loci than selected loci (Fig. 3g-i).

3.3.2 Effect of the selection regime on both simulated traits

While little variation was observed between different locus types within a selection regime, the average divergence was relatively higher when both traits were under parallel selection, compared to when both traits were under divergent selection (Fig. 1a-b,2a-b). Intermediate levels of divergence were observed when one trait was under parallel selection and the other trait was under divergent selection. For simulations without DMIs, a paired t-test indicated that population size was significantly larger for simulations under divergent selection ($M= 65.74$, $SD= 8.63$) than simulations under parallel selection ($M=51.44$, $SD= 6.96$) ; $t(99)= 13.63$, $p \leq 2.2 \times 10^{-16}$.

HMM analysis showed little variation across the different selection regimes after 1000 generations (Fig. 3). Across all three locus types, similar percentages of island and valley formation were observed between loci of the same type under different selection regimes (Fig. 3). Comparisons between trait 1 and trait 2 for all three selection regimes did not show a consistent pattern in island and valley frequency (Fig. 3).

3.3.3 Effect of migration on the genomic landscape

We showed that migration rate influenced the overall level of divergence (Fig. 2,S1). After 1000 generations with a medium migration rate ($m_{ij} = 0.01$) in the absence of DMIs, genomic islands were not observed across the average landscape of divergence of the 100 independent simulations (Fig. 2). However, as shown by the HMM, islands and valleys

were evenly distributed across the genomic landscape over the 100 independent simulations (Fig. 3). Under a lower migration rate ($m_{ij} = 0.0001$), across the average landscape of divergence of the 100 independent simulations, genomic islands were observed at selected loci for simulations without DMIs (Fig. S1). This was reflected by the HMM, which showed that the majority of islands formed at selected loci, while valleys were evenly distributed among selected and neutral loci (Fig. S2).

Under a low migration rate, simulations without DMIs showed a similar level of divergence to simulations with a high DMI probability under a medium migration rate (Fig. 2,S1). Furthermore, the application of a high DMI probability did not influence the landscape of divergence under a low migration rate (Fig. S1,S3).

3.4. Discussion

We extended an IBM approach to explore how DMIs, with and without gene flow, influence the genomic landscape of divergence under parallel and divergent selection. Using parameters and data based on empirical data from an island-colonising bird population (Sendell-Price et al., 2020), we show that DMIs at neutral loci and loci under parallel or divergent selection are able to generate divergence, despite gene flow. By adding an additional layer of complexity when analysing two traits concurrently affected by similar (both parallel or both divergent selection) or different selection regimes (one trait divergent and the other parallel selection), we highlight that incorporating different levels of complexity is key in explaining the formation of specific patterns in the genomic landscape of divergence, as also shown by Ali et al. (2023) (Chapter 2).

3.4.1 DMIs influence the level of divergence between populations

Our model provides insights into the interlinked processes of DMI accumulation, gene flow and the level of genomic divergence between diverging populations. In our model, we observed that DMIs can lead to genomic divergence despite medium levels of migration, for both parallel and divergent selection. While the formation of DMIs under divergent selection has been theorised, the formation of DMIs under parallel selection with gene flow is challenging (Schluter, 2009). Under parallel selection, mutations are adapted globally i.e. adapted to both populations (in contrast to locally with divergent selection). Thus, the globally adapted mutations, that generate negative epistatic interactions, are removed by selection (Schluter, 2009). On the other hand, locally adapted mutations are maintained, due to divergent selection generating a barrier to gene flow (Rundle & Nosil, 2005). Our simulations of divergent selection with gene flow reflect this with the observed increase in F_{ST} with divergent selection. Thus, our simulations add to the body of empirical studies highlighting DMIs can elevate divergence under divergent selection with gene flow (Anderson & Harmon, 2014; Bank et al., 2012; Blanckaert & Hermisson, 2018). While DMI-mediated divergence is challenging under parallel selection with gene flow, under certain conditions it can persist. As demonstrated by Nosil & Flaxman (2011), if the DMIs have a similar fitness (as with parallel selection), divergence can occur with medium levels of gene flow; although, a caveat of low gene flow is necessary, in the case of the fixation of a DMI allele. In this case, the frequency of the allele across the populations dictates the limit of gene flow (Höllinger & Hermisson, 2017; Nosil & Flaxman, 2011). In our simulations, we develop the aforementioned studies of DMI formation by simulating multiple loci additively, in contrast to either Fisher's geometric model-based approaches or simulations of a few loci. Therefore, the fixation of alleles is less frequent, as multiple routes exist to

a phenotypic optima in the *glads* model (Ali et al., 2023). At these polymorphic sites, DMIs may form facilitating divergence, under parallel selection with gene flow. By using the *glads* model, we also support the findings of Anderson & Harmon (2014), who provide evidence for reproductive isolation with DMIs under gene flow (under parallel and divergent selection), although we also demonstrate this in a model mimicking biological systems (with the potential for parameterisation to a specific biological system).

A key difference between our approach and the commonly used Fisher's geometric model (Fraïsse et al., 2016; Orr, 2005; Yamaguchi & Otto, 2020), is that we simulated DMIs at neutral loci as well as loci under selection. Additionally, the basis of Fisher's geometric model is to explore how the number of traits affects the fitness of an individual in contrast to the genomic landscape. Thus, these models do not easily facilitate simulating heterogeneous combinations of microevolutionary processes and their effect on the genomic landscape. Under gene flow, DMIs have been theorised to fail to form at neutral loci (Coyne & Orr, 2004a). This is because the mutation may be neutral in its population of origin, but in a different genetic environment it is subjected to negative epistatic interactions, and thus is purged by selection (Bank et al., 2012). As discussed above, this is predicated on the gene flow limit being dependent on allele frequency (i.e. in the case of allele fixation, the gene flow rate must be negligible) (Nosil & Flaxman, 2011). In our simulation, the formation of DMIs and allele frequency of neutral alleles is stochastic and thus at polymorphic neutral loci, DMIs can exist, despite medium levels of gene flow. As discussed above, DMIs under parallel selection and at neutral loci may be maintained under medium levels of gene flow (Nosil & Flaxman, 2011). In our model, the medium migration rate we simulate is the upper limit of medium gene flow under the parameter definition of Nosil & Flaxman (2011). Additionally, in our simulations we

simulate migration rate, which is likely to be higher than the gene flow rate, as not all migrant individuals will breed. Thus, our simulations fall into the parameter space suggested to facilitate divergence by DMI at neutral loci and loci under parallel selection (Nosil & Flaxman, 2011).

3.4.2 The level of divergence could be counterbalanced with the concurrent effect of different types of selection

The different level of divergence observed for parallel and divergent could be explained by the smaller population sizes reached under parallel selection, in which differentiation between populations accumulates faster (Relethford, 1991). In our simulations, when the traits were under parallel selection, migrant individuals were adapted globally, and could perform better in the new population than when migrants were locally adapted i.e. when the traits were under divergent selection. However, the opposite pattern was possible in some of the independent simulation, and in simulations without migration (Quilodrán, Ruegg, et al., 2020a).

In this study, our simulations highlight the complexity of predicting the genomic landscape of divergence, as the concurrent effects of selection on different traits influences the genomic landscape. For instance, divergent selection typically generates genomic valleys, but with the concurrent occurrence of positive frequency dependent selection, genomic islands form (Ali et al., 2023). As simulated in this study, different types of selection, such as divergent and parallel selection, can occur concurrently in a population on different traits (Langerhans & Riesch, 2013). This highlights the need for caution when interpreting patterns in the genomic landscape, as divergence levels may be reduced (as shown here) or the reverse pattern generated, as a consequence of additional layers of complexity.

3.5. Conclusion

We developed a modelling approach to explore the concurrent effect of both parallel selection and divergent selection on multiple traits, which provides insights into how divergence and reproductive isolation form. By exploring the interplay between gene flow, population size, type of selection, and genetic structure of two traits on the genomic landscape of divergence from independent simulations, we show that 1) under gene flow, divergence can be facilitated by DMIs occurring at neutral loci and loci under parallel or divergent selection, and 2) that variable outputs may be expected when different layers of complexity are concurrently analysed. While DMIs provide a mechanism for the generation and maintenance of divergence under medium levels of migration, the divergence along the genome is influenced by the different evolutionary processes concurrently affecting a given biological system. Our results also support previous research highlighting that caution is recommended in inferring the causal evolutionary processes of genomic landscapes in empirical studies (Ali et al., 2023; Funk et al., 2021; Quilodr  n, Ruegg, et al., 2020a). Our quantitative framework is useful for studying how DMIs move through populations over time under the influence of different evolutionary processes, including stochasticity and selection on multiple traits. Methods such as approximate Bayesian computation and machine learning, can be used to generate summary statistics across a range of different parameters. From these summary statistics, parameters can be accepted and rejected in accordance with the observed empirical data (Csill  ry et al., 2010). Thus, the range of potential drivers that generate the empirically observed genomic landscape for a particular taxon can potentially be narrowed.

3.6. References

- Ali, H. A. A., Coulson, T., Clegg, S. M., & Quilodrán, C. S. (2023). The effect of divergent and parallel selection on the genomic landscape of divergence. *Molecular Ecology, Advance online publication*. <https://doi.org/https://doi.org/10.1111/mec.17225>
- Anderson, C. J. R., & Harmon, L. (2014). Ecological and mutation-order speciation in digital organisms. *American Naturalist*, 183(2), 257–268. <https://doi.org/10.1086/674359>
- Anderson, E. C., Quilodran, C. S., & Coulson, T. (2019). eriqande/glads: Companion to MIEE article, Quilodran et al. (Version v1.0). *Zenodo*. <https://doi.org/https://doi.org/10.5281/zenodo.3516224>
- Andrello, M., & Manel, S. (2015). MetaPopGen: An R package to simulate population genetics in large size metapopulations. *Molecular Ecology Resources*, 15(5), 1153–1162. <https://doi.org/10.1111/1755-0998.12371>
- Bank, C., Bürger, R., & Hermisson, J. (2012). The limits to parapatric speciation: Dobzhansky-Muller incompatibilities in a continent-Island model. *Genetics*, 191(3), 845–863. <https://doi.org/10.1534/genetics.111.137513>
- Blanckaert, A., & Hermisson, J. (2018). The limits to parapatric speciation II: Strengthening a preexisting genetic barrier to gene flow in parapatry. *Genetics*, 209(1), 241–254. <https://doi.org/10.1534/genetics.117.300652>
- Brook, B. W., & Kikkawa, J. (1998). Examining threats faced by island birds: A population viability analysis on the Capricorn silveryeye using long-term data. *Journal of Applied Ecology*, 35(4), 491–503. <https://doi.org/10.1046/j.1365-2664.1998.3540491.x>
- Campbell, C. R., Poelstra, J. W., & Yoder, A. D. (2018). What is Speciation Genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species.

Biological Journal of the Linnean Society, 124(4), 561–583.

<https://doi.org/10.1093/biolinnean/bly063>

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. <https://doi.org/10.1017/S0016672397002954>

Chevin, L. M., Decorzent, G., & Lenormand, T. (2014). Niche dimensionality and the genetics of ecological speciation. *Evolution*, 68(5), 1244–1256. <https://doi.org/10.1111/evo.12346>

Chou, J.-Y., Hsu, P.-C., & Leu, J.-Y. (2022). Enforcement of Postzygotic Species Boundaries in the Fungal Kingdom. *Microbiology and Molecular Biology Reviews*. <https://doi.org/10.1128/membr.00098-22>

Clegg, S. M., Degnan, S. M., Kikkawa, J., Moritz, C., Estoup, A., & Owens, I. P. F. (2002). Genetic consequences of sequential founder events by an island-colonizing bird. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12), 8127–8132. <https://doi.org/10.1073/pnas.102583399>

Clegg, S. M., Degnan, S. M., Moritz, C., Estoup, A., Kikkawa, J., & Owens, I. P. F. (2002). Microevolution in island forms: The roles of drift and directional selection in morphological divergence of a passerine bird. *Evolution*, 56(10), 2090–2099. <https://doi.org/10.1111/j.0014-3820.2002.tb00134.x>

Clegg, S. M., Frentiu, F. D., Kikkawa, J., Tavecchia, G., & Owens, I. P. F. (2008). 4000 Years of phenotypic change in an island bird: Heterogeneity of selection over three microevolutionary timescales. *Evolution*, 62(9), 2393–2410. <https://doi.org/10.1111/j.1558-5646.2008.00437.x>

Clegg, S. M., & Phillimore, A. B. (2010). The influence of gene flow and drift on genetic and phenotypic divergence in two species of *Zosterops* in Vanuatu. *Philosophical*

- Transactions of the Royal Society B: Biological Sciences*, 365(1543), 1077–1092.
<https://doi.org/10.1098/rstb.2009.0281>
- Corbett-Detig, R. B., Zhou, J., Clark, A. G., Hartl, D. L., & Ayroles, J. F. (2013). Genetic incompatibilities are widespread within species. *Nature*, 504(7478), 135–137.
<https://doi.org/10.1038/nature12678>
- Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sinauer Associates, Inc. Publishers.
- Csilléry, K., Blum, M. G. B., Gaggiotti, O. E., & François, O. (2010). Approximate Bayesian Computation (ABC) in practice. In *Trends in Ecology and Evolution* (Vol. 25, Issue 7, pp. 410–418). <https://doi.org/10.1016/j.tree.2010.04.001>
- Curat, M., Gerbault, P., Di, D., Nunes, J. M., & Sanchez-Mazas, A. (2015). Forward-in-time, spatially explicit modeling software to simulate genetic lineages under selection. *Evolutionary Bioinformatics*, 11, 27–39. <https://doi.org/10.4137/EBO.S33488>
- Feder, J. L., Gejji, R., Yeaman, S., & Nosil, P. (2012). Establishment of new mutations under divergence and genome hitchhiking. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1587), 461–474. <https://doi.org/10.1098/rstb.2011.0256>
- Feder, J. L., & Nosil, P. (2009). Chromosomal inversions and species differences: When are genes affecting adaptive divergence and reproductive isolation expected to reside within inversions? *Evolution*, 63(12), 3061–3075. <https://doi.org/10.1111/j.1558-5646.2009.00786.x>
- Feder, J. L., & Nosil, P. (2010). The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution: International Journal of Organic Evolution*, 64(6), 1729–1747. <https://doi.org/10.1111/j.1558-5646.2009.00943.x>
- Fisher, R. A. (1930). The genetical theory of natural selection. In *The genetical theory of natural selection*. Clarendon Press. <https://doi.org/10.5962/bhl.title.27468>

- Fraïsse, C., Gunnarsson, P. A., Roze, D., Bierne, N., & Welch, J. J. (2016). The genetics of speciation: Insights from Fisher's geometric model. *Evolution; International Journal of Organic Evolution*, 70(7), 1450–1464. <https://doi.org/10.1111/evo.12968>
- Funk, E. R., Spellman, G. M., Winker, K., Withrow, J. J., Ruegg, K. C., Zavaleta, E., & Taylor, S. A. (2021). Phylogenomic Data Reveal Widespread Introgression Across the Range of an Alpine and Arctic Specialist. *Systematic Biology*, 70(3), 527–541. <https://doi.org/10.1093/sysbio/syaa071>
- Gavrilets, S. (2014). Models of speciation: Where are we now? *Journal of Heredity*, 105(S1), 743–755. <https://doi.org/10.1093/jhered/esu045>
- Harte, D. (2017). *HiddenMarkov: Hidden Markov Models* (R package version 1.8-11). Statistics Research Associates. <http://www.statsresearch.co.nz/dsh/sslib/>
- Hoban, S., Bertorelle, G., & Gaggiotti, O. E. (2012). Computer simulations: Tools for population and evolutionary genetics. *Nature Reviews Genetics*, 13(2), 110–122. <https://doi.org/10.1038/nrg3130>
- Höllinger, I., & Hermisson, J. (2017). Bounds to parapatric speciation: A Dobzhansky–Muller incompatibility model involving autosomes, X chromosomes, and mitochondria. *Evolution*, 71(5), 1366–1380. <https://doi.org/10.1111/evo.13223>
- Kawakami, T., Smeds, L., Backström, N., Husby, A., Qvarnström, A., Mugal, C. F., Olason, P., & Ellegren, H. (2014). A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. *Molecular Ecology*, 23(16), 4035–4058. <https://doi.org/10.1111/mec.12810>
- Kikkawa, J., & Wilson, J. M. (1983). Breeding and dominance among the Heron Island Silvereyes *Zosterops lateralis chlorocephala*. *Emu-Austral Ornithology*, 83(3), 181–198.

- Langerhans, R. B., & Riesch, R. (2013). Speciation by selection: A framework for understanding ecology's role in speciation. *Current Zoology*, 59(1), 31–52. <https://doi.org/10.1093/czoolo/59.1.31>
- Marques, D. A., Lucek, K., Meier, J. I., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2016). Genomics of Rapid Incipient Speciation in Sympatric Threespine Stickleback. *PLoS Genetics*, 12(2), 1–34. <https://doi.org/10.1371/journal.pgen.1005887>
- Martin, G., & Lenormand, T. (2006). a General Multivariate Extension of Fisher's Geometrical Model and the Distribution of Mutation Fitness Effects Across Species. *Evolution*, 60(5), 893. <https://doi.org/10.1554/05-412.1>
- Matute, D. R., Butler, I. A., Turissini, D. A., & Coyne, J. A. (2010). A test of the snowball theory for the rate of evolution of hybrid incompatibilities. *Science*, 329(5998), 1518–1521.
- Michel, A. P., Sim, S., Powell, T. H. Q., Taylor, M. S., Nosil, P., & Feder, J. L. (2010). Widespread genomic divergence during sympatric speciation. *Proceedings of the National Academy of Sciences of the United States of America*, 107(21), 9724–9729. <https://doi.org/10.1073/pnas.1000939107>
- Nosil, P., & Flaxman, S. M. (2011). Conditions for mutation-order speciation. *Proceedings of the Royal Society B: Biological Sciences*, 278(1704), 399–407. <https://doi.org/10.1098/rspb.2010.1215>
- Orr, H. A. (2005). The genetic theory of adaptation: A brief history. *Nature Reviews Genetics*, 6(2), 119–127. <https://doi.org/10.1038/nrg1523>
- Paradis, E. (2010). Pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics*, 26(3), 419–420. <https://doi.org/10.1093/bioinformatics/btp696>
- Quilodrán, C. S., Ruegg, K., Sendell-Price, A. T., Anderson, E. C., Coulson, T., & Clegg, S. M. (2020). The multiple population genetic and demographic routes to islands of genomic

- divergence. *Methods in Ecology and Evolution*, 11(1), 6–21.
<https://doi.org/10.1111/2041-210X.13324>
- R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.r-project.org/>
- Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., Noor, M. A. F., Mehlig, B., & Westram, A. M. (2017). Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, 30(8), 1450–1477. <https://doi.org/10.1111/jeb.13047>
- Relethford, J. H. (1991). Effect of changes in population size on genetic microdifferentiation. *Human Biology*, 629–641.
- Rundle, H. D., & Nosil, P. (2005). Ecological speciation. *Ecology Letters*, 8(3), 336–352.
<https://doi.org/10.1111/j.1461-0248.2004.00715.x>
- Schluter, D. (2009). Evidence for ecological speciation and its alternative. *Science*, 323(5915), 737–741. <https://doi.org/10.1126/science.1160006>
- Schluter, D., & Conte, G. L. (2009). Genetics and ecological speciation. *Proceedings of the National Academy of Sciences of the United States of America*, 106 Suppl, 9955–9962.
<https://doi.org/10.1073/pnas.0901264106>
- Sedghifar, A., Brandvain, Y., & Ralph, P. (2016). Beyond clines: lineages and haplotype blocks in hybrid zones. *Molecular Ecology*, 25(11), 2559–2576.
<https://doi.org/10.1111/mec.13677>
- Sendell-Price, A. T., Ruegg, K. C., Anderson, E. C., Quilodrán, C. S., Van Doren, B. M., Underwood, V. L., Coulson, T., & Clegg, S. M. (2020). The Genomic Landscape of Divergence Across the Speciation Continuum in Island-Colonising Silvereyes (*Zosterops lateralis*). *G3: Genes, Genomes, Genetics*, 10(9), 3147–3163.
<https://doi.org/10.1534/g3.120.401352>

- Simon, A., Bierne, N., & Welch, J. J. (2018). Coadapted genomes and selection on hybrids: Fisher's geometric model explains a variety of empirical patterns. *Evolution Letters*, 2(5), 472–498. <https://doi.org/10.1002/evl3.66>
- Turelli, M., & Orr, H. A. (2000). Dominance, epistasis and the genetics of postzygotic isolation. *Genetics*, 154(4), 1663–1679. <https://doi.org/10.1093/genetics/154.4.1663>
- Waxman, D., & Welch, J. J. (2005). *Fisher's and Haldane's Microscope Ellipse*. 166(4), 447–457.
- Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, 38(6), 1358–1370. <https://doi.org/10.2307/2408641>
- Wolf, J. B. W., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, 18(2), 87–100. <https://doi.org/10.1038/nrg.2016.133>
- Yamaguchi, R., & Iwasa, Y. (2017). A tipping point in parapatric speciation. *Journal of Theoretical Biology*, 421, 81–92. <https://doi.org/10.1016/j.jtbi.2017.03.018>
- Yamaguchi, R., & Otto, S. P. (2020). Insights from Fisher's geometric model on the likelihood of speciation under different histories of environmental change. *Evolution*, 74(8), 1603–1619. <https://doi.org/10.1111/evo.14032>

3.7. Figures

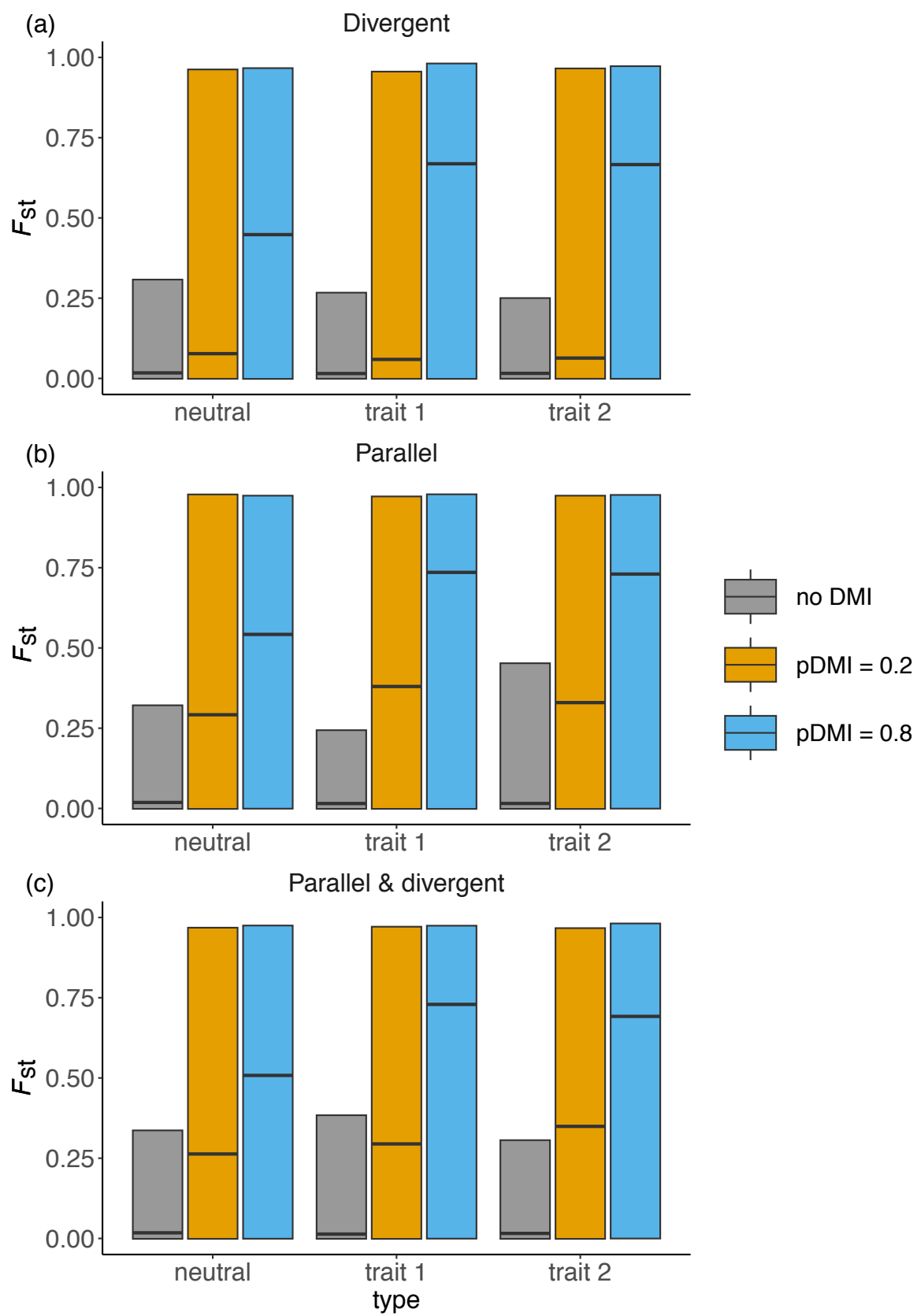


Figure 1. Mean F_{ST} histogram for the neutral and selected loci. The F_{ST} values were computed after 1000 generations for 100 independent simulations under a medium migration rate ($m_{ij} = 0.01$). The boxes represent the minimum and maximum range, and the bar represents the median. Three different DMI probabilities are compared ($p_{DMI} = 0$ & 0.8 & 0.2) under three different selection regimes: (a) both selected traits are under divergent selection; (b) both selected traits are under parallel selection; (c) trait 1 is under parallel selection and trait 2 is under diverget selection.

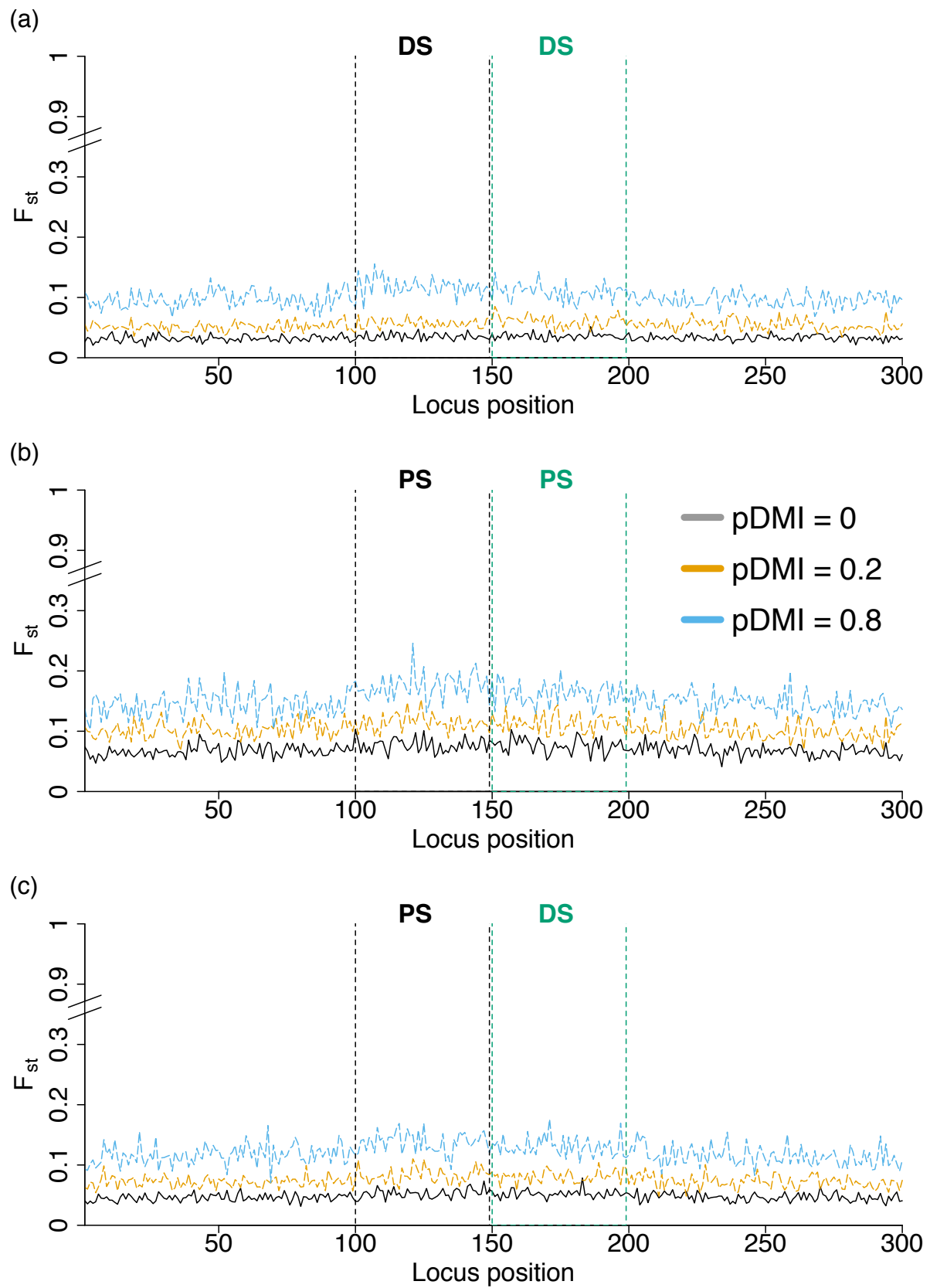


Figure 2. The genomic landscape of divergence between two populations. The landscape is illustrated by the average F_{ST} from 100 simulations after 1000 generations under a medium migration rate ($m_{ij} = 0.01$). Vertical dashed black lines represent loci contributing to trait 1 and vertical dashed green lines those for trait 2. Three DMI mutation probabilities (p_{DMI}) are compared where (a) both traits are under divergent selection (DS); (b) both traits are under parallel selection (PS); and (c) trait 1 is under parallel selection and trait 2 is under divergent selection.

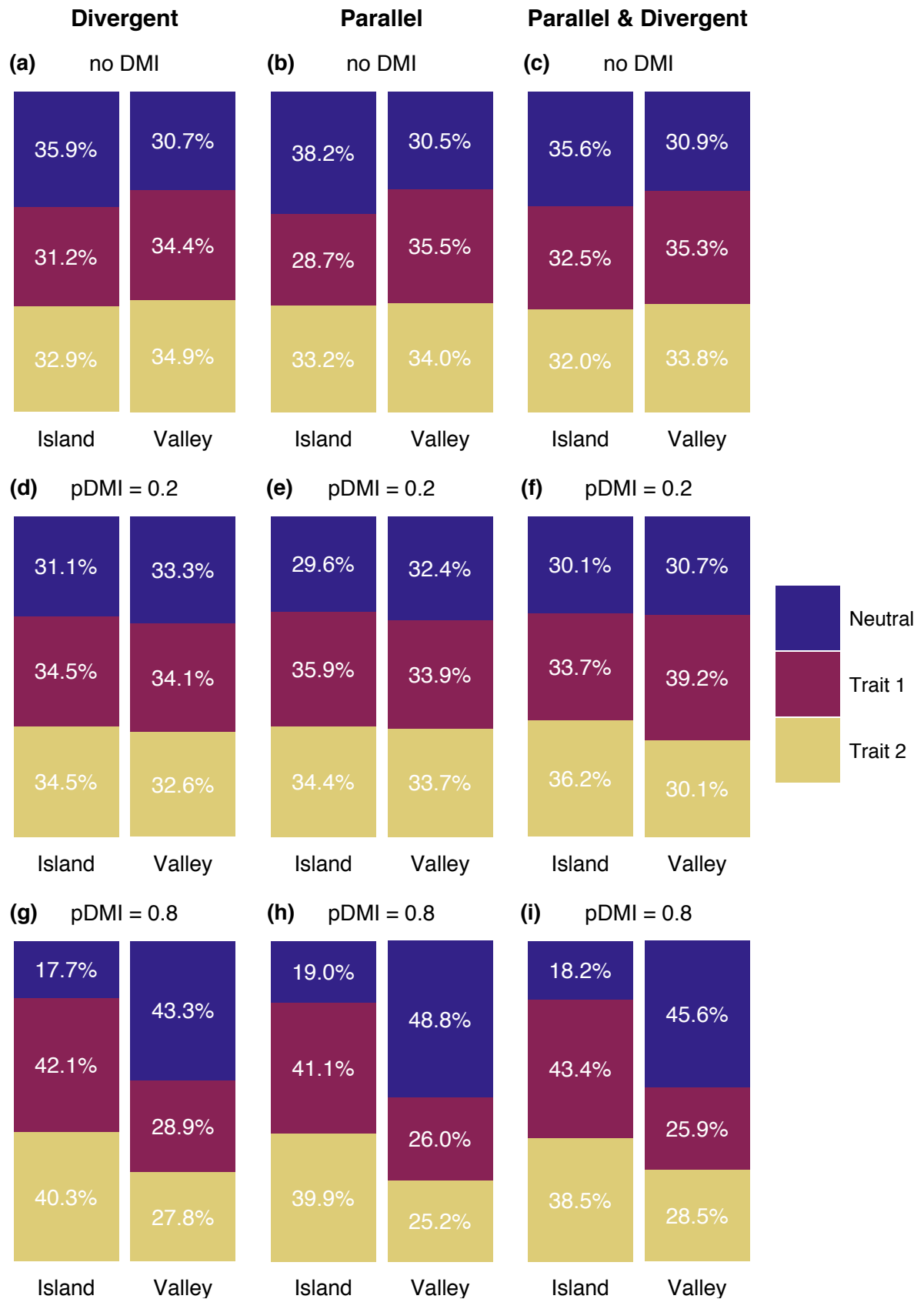


Figure 3. Proportion of genomic islands and valleys observed using the Hidden Markov Model. The proportions are observed for 100 simulations of 1000 generations each, and a medium migration rate ($m_{ij} = 0.01$). The proportions are calculated for three different DMI probabilities ($p_{DMI} = 0$ & 0.2 & 0.8) and under three different selection regimes: (left column) both selected traits are under divergent selection; (middle column) both selected traits are under parallel selection; (right column) trait 1 is under parallel selection and trait 2 is under divergent selection.

3.8. Supporting information

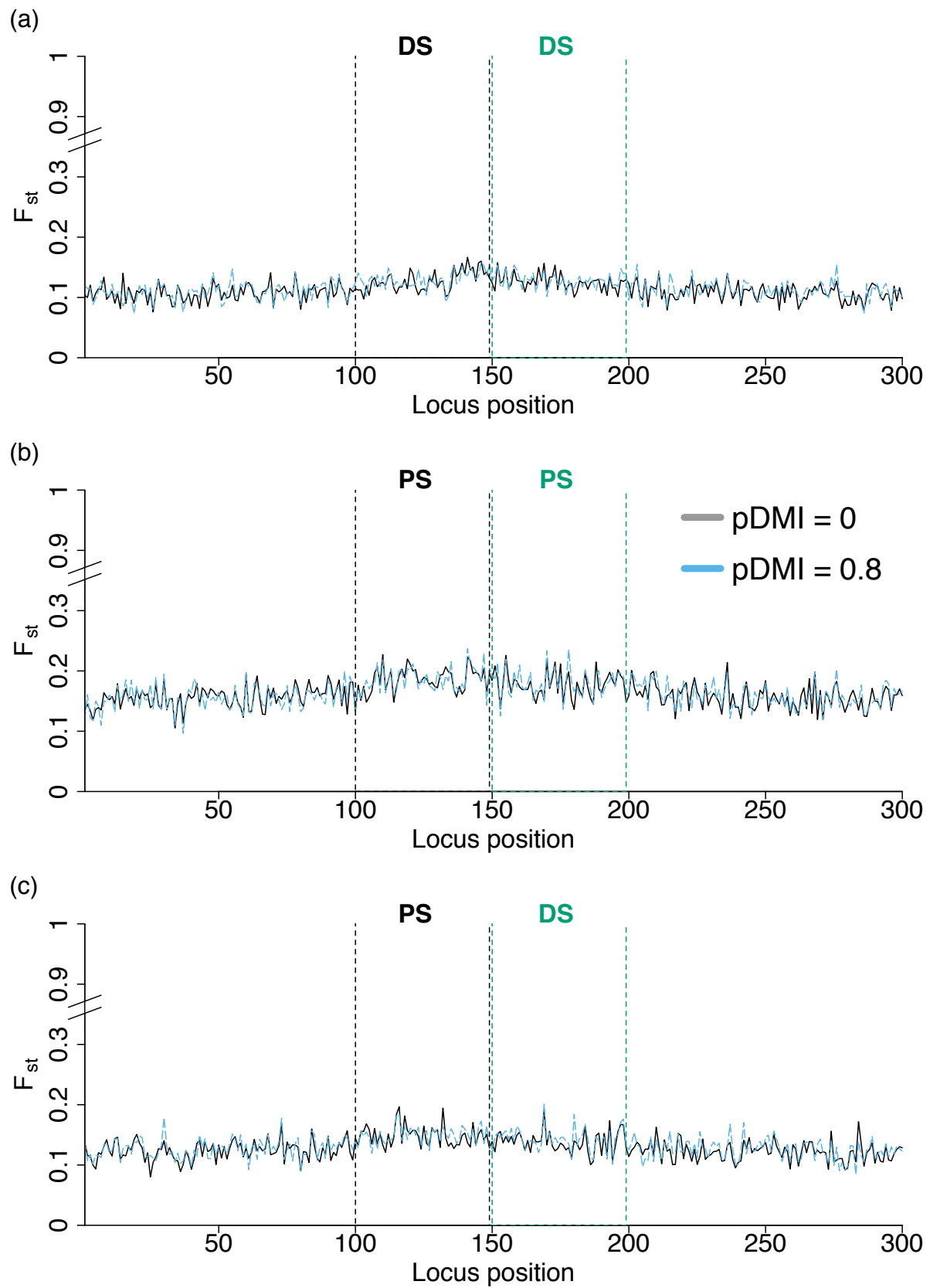


Figure S1. The genomic landscape of divergence between two populations. The landscape is illustrated by the average F_{ST} from 100 simulations after 1000 generations under a low migration rate ($m_{ij} = 0.0001$). Vertical dashed black lines represent loci contributing to trait 1 and vertical dashed green lines those for trait 2. Two DMI mutation probabilities (p_{DMI}) are compared where (a) both traits are under divergent selection (DS); (b) both traits are under parallel selection (PS); and (c) trait 1 is under parallel selection and trait 2 is under divergent selection.

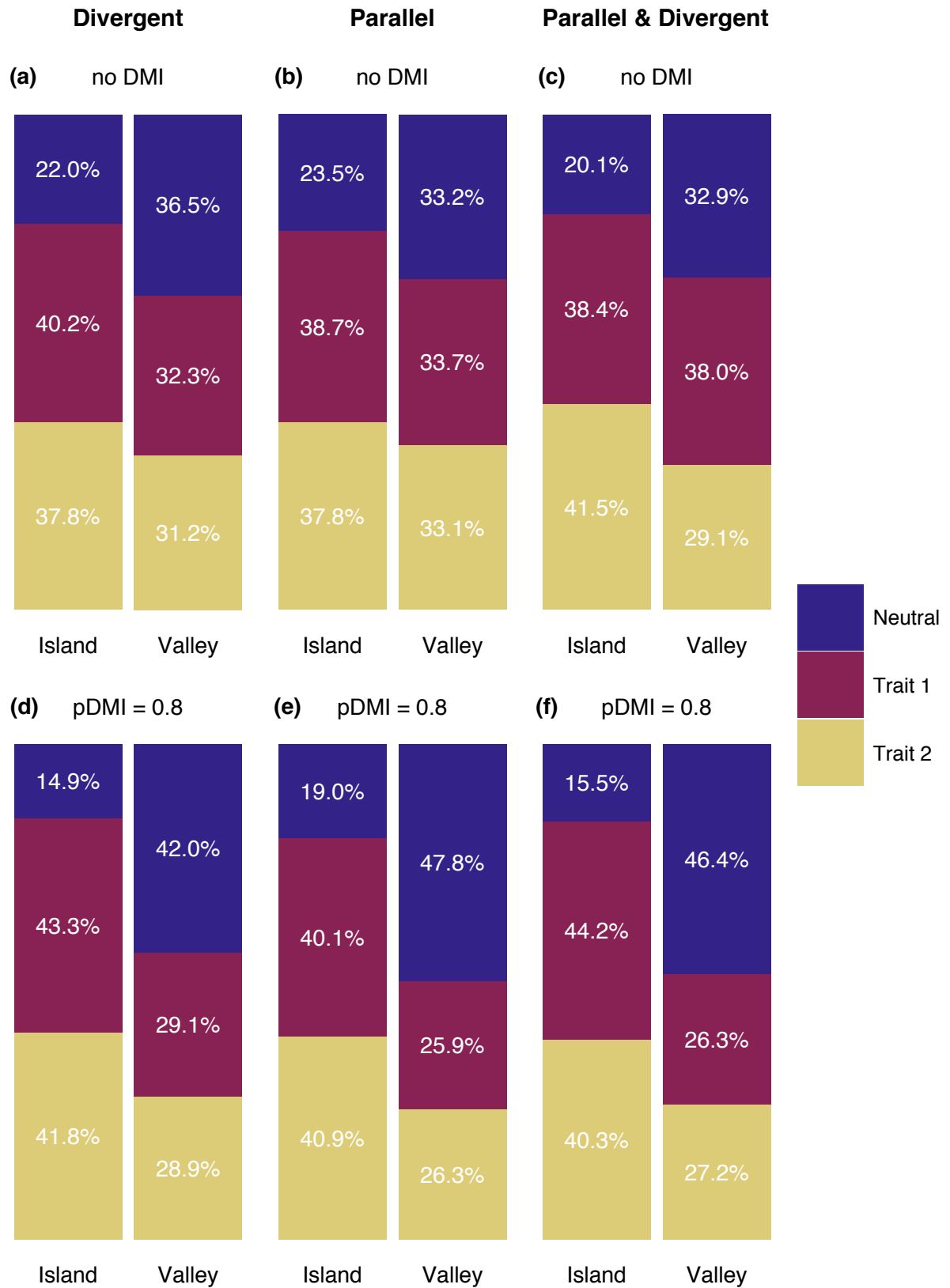


Figure S2. Proportion of genomic islands and valleys observed using the Hidden Markov Model. The proportions are observed for 100 simulations of 1000 generations each, and a low migration rate ($m_{ij} = 0.0001$). The proportions are calculated for two different DMI probabilities ($p_{DMI} = 0$ & 0.8) and under three different selection regimes: (left column) both selected traits are under divergent selection; (middle column) both selected traits are under parallel selection; (right column) trait 1 is under parallel selection and trait 2 is under divergent selection.

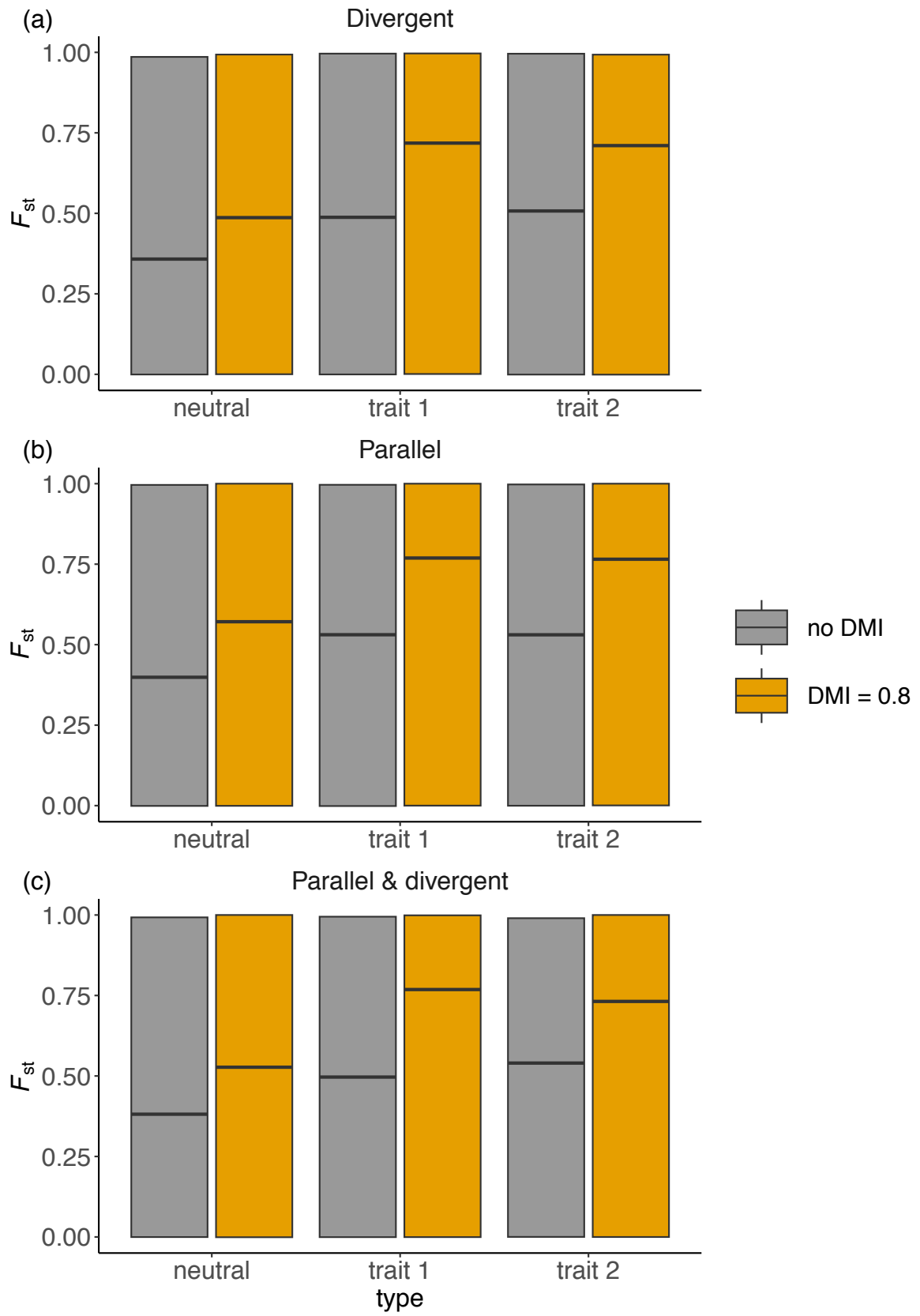


Figure S3. Mean F_{ST} histogram for the neutral and selected loci. The F_{ST} values were computed after 1000 generations for 100 independent simulations at a low migration rate ($m_{ij} = 0.0001$). The boxes represent the minimum and maximum range, and the bar represents the median. Two different DMI probabilities are compared ($p_{DMI} = 0$ & 0.8) under three different selection regimes: (a) both selected traits are under divergent selection; (b) both selected traits are under parallel selection; (c) trait 1 is under parallel selection and trait 2 is under divergent selection.

4. Effects of selection and gene flow barriers on the genomic landscape of introgression

Abstract

Many species hybridise with a closely related species, and with advances in genomic sequencing, we now know that it is far more common than previously thought. Hybridisation often results in a heterogeneous genomic landscape of introgression, which is characterised by regions of high and low levels of DNA transferred across the genomes of divergent lineages. A key question is how different evolutionary processes shape the genomic landscape of introgression, such as selection generating adaptive and maladaptive introgression. Here, we extend an individual-based model to understand the formation of heterogeneous genomic landscapes of introgression and examine the concurrent effects of a gene flow barrier in hybrids, selection, genes of large effect, recombination, and drift. While independent replicates with the same settings show considerable variation in levels of introgression after 2000 number of generations, our results showed that introgression under positive and negative selection is frequently associated with elevated and reduced rates of introgression, respectively. We also show that these elevated and reduced levels of introgression are amplified when we simulate hybrids with a gene flow barrier, e.g. the simulation of Haldane's rule, where hybrids of the heterogametic sex have a lower fitness than the homogametic sex. We highlight that the expected effect on the genomic landscape of introgression of a single evolutionary process considered in isolation may be different from that expected when evolutionary processes interact. Our framework provides insight into how different evolutionary

processes shape the genomic landscape of introgression, and offers the opportunity to narrow down the range of processes that underlie empirical patterns.

Keywords: adaptive introgression, genomic landscape of introgression, *glads* model, Haldane's rule, hybridisation barriers, recombination, selection

4.1. Introduction

Hybridisation is more common than previously thought, with at least 10% of mammals and 25% of plants currently hybridising with a closely related species (Mallet, 2005). Hybridisation can influence evolution in different ways, ranging from species differentiation to species fusion (Seehausen, 2004). While hybridisation is recognized as playing a major role in evolution, the consequences on particular evolutionary trajectories is poorly understood, especially concerning the genomic regions influenced by hybridisation and adaptive introgression (Adavoudi & Pilot, 2022; Edelman & Mallet, 2021).

Introgression describes the transfer of DNA between genetically distinct lineages (i.e., subspecies and species) as a consequence of hybridisation. An introgressed allele is considered adaptive when it has a beneficial effect on fitness; and maladaptive when it has a fitness disadvantage. While the vast majority of introgression events are maladaptive (Veller et al., 2023), there are various examples of adaptive introgression that include insecticide resistance in *Anopheles* mosquitos and seasonal camouflage in *Lepus* hares (Jones et al., 2018; Norris et al., 2015; Taylor & Larson, 2019). However, adaptive introgression is difficult to empirically demonstrate, as the majority of well understood examples come from single large effect loci that either have a clear adaptive

function or arise from recent introgression events (Taylor & Larson, 2019). In contrast to focusing on simple traits that are known to be adaptive, more recently studies have taken to identifying the relationship between evolutionary processes and genome-wide introgression patterns to help identify adaptive introgression (S. H. Martin & Jiggins, 2017).

The genomic landscape of introgression (GLI) describes the heterogeneous variation in introgression levels across the genomes. It has two main topographic features: islands and valleys, which refer to areas of high and low levels of introgression relative to the background level of introgression, respectively. However, interpreting the effects of evolutionary processes on the GLI is challenging, as the effect of an evolutionary process can change with the concurrent influence of another evolutionary process, as well as with the architecture of genomes (Duranton & Pool, 2022). For example, under negative selection, introgression levels typically positively correlate with recombination rates (Ravinet et al., 2018), while under positive selection a negative correlation is typically observed (Duranton & Pool, 2022). These correlations can be reversed under different evolutionary circumstances, such as particular Dobzhansky Muller incompatibility dominance hierarchies (different fitness impacts of a Dobzhansky Muller incompatibility), or when both populations experience substantial genetic load, in which case introgression levels could negatively correlate with recombination under negative selection (Duranton & Pool, 2022). Furthermore, levels of introgression can be reduced in the heterogametic sex as a consequence of its reduced fitness (Haldane's rule) and increased action of Dobzhansky Muller incompatibilities (Ottenburghs, 2022). The sex chromosome of the heterogametic sex of a species also has a lower effective population size, thereby exposing it to larger effects of drift (which negatively correlates with levels of introgression) (Ottenburghs, 2022).

Modelling approaches have been used to explore the effect of different evolutionary processes on the genomic landscape of divergence, the heterogeneous distribution of differentiation between diverging lineages, and can be used as a base for developing models to study the GLI. The various approaches have provided insights into the effect of recombination on negative selection purging deleterious alleles, although these approaches were limited to small numbers of loci (Barton & Bengtsson, 1986; Barton, 1983). More recently, modelling approaches have been able to simulate genomic data (Haller et al., 2019; Haller & Messer, 2019), focused on the effects of selection and recombination (Dagilis & Matute, 2023; Duranton & Pool, 2022; B. Y. Kim et al., 2018; Veller et al., 2023). However, these approaches are often limited to uniform distributions of loci, providing limited insights into the effects of selection on small populations. Here, we extended the “*glads*” (genomic landscape of divergence simulations) individual-based model (Quilodrán, Ruegg, et al., 2020b), expanding on the effects of positive and negative selection and variable genomic architectures (including non-uniform distributions of loci) on the GLI. We made use of the flexible fitness function of this forward-in-time models to explore the effects of selection i.e., positive, negative, and equivalent (where the fitness of the selected loci in both populations is the same) introgression, and their interaction with other evolutionary processes. Additionally, we implemented a non-recombining sex chromosome and a sex-specific reduced hybrid gene flow barrier, to simulate i) Haldane’s rule, where only the fitness of the heterogametic sex is reduced; and ii) a gene flow barrier for hybrids, where the fitness of both sexes is reduced. Our aim is to better understand the formation of the heterogeneous GLI, providing a framework for exploring the patterns that may emerge with particular combinations of evolutionary processes.

4.2. Methods

4.2.1 Description of the model

Simulations were performed using the individual-based modelling framework ‘glads’ (*Genomic Landscape of Divergence Simulations*) (Quilodrán, Ruegg, et al., 2020b). The purpose of this approach is to model the impact of different evolutionary processes on the genomic landscape of divergence for two diverging lineages. Our motivation was to adapt this framework to examine how different evolutionary processes impact the GLI.

This approach simulates the divergence of two or more populations (diverging lineages) across three hierarchical levels: genotypes, phenotypic traits and demographic rates. This means that the population dynamics and distribution of genotypes and phenotypic traits are emerging properties of the model. Two lineages composed of genetically and sexually defined diploid individuals were simulated. The alleles at each locus under selection (n_a) contribute breeding values (b_v) additively alongside a stochastic environmental component (ε_{env}) to give the phenotype (z) of the individual (eq.1).

$$z = \sum_{v=1}^{n_a} b_v + \varepsilon_{env}(0, \sigma_{env}) \quad (1)$$

The fitness (ω) describes the potential reproductive output of the individual, and is calculated by applying the phenotype of that individual to a modified Gaussian function (eq.2). This function is parameterised by a maximum fitness (b_0), which is achieved when an individual has the optimal phenotype (b_1), and the variance of the Gaussian curve (b_2). By setting different values of b_1 between the populations (or lineages), divergent selection

can be simulated as the fitness of a phenotype differs between them. Similarly, parallel selection is introduced by setting the same value of b_1 in both populations. The fitness function is also additively regulated by a density dependence factor (b_3) that acts on the population size (N) to avoid exponential growth, as well as a stochastic demographic factor (ε_{dem}).

$$\omega = b_0 e^{\frac{1}{2}(\frac{4z-b_1n_a}{b_2n_a})^2} - b_3N + \varepsilon_{dem}(0, \sigma_{dem}) \quad (2)$$

As the model proceeds in discrete time steps (t), representing generations, breeding pairs form regulated by the number of females. During reproduction, recombination points are determined between the parents using the recombination map parameter, a vector of recombination rates between adjacent loci (ρ), where $\rho = 0.5$ represents independent assortment between neighbouring loci, a value of $\rho < 0.5$ represents linkage, and a value of $\rho = 0$ represents no recombination. New mutations are incorporated at a given rate per generation (μ). The sex of the offspring is assigned by a within-population sex ratio. The new generation of offspring have the potential to migrate at the rate given by the migration rate parameter (m_{ij}), which is the probability of a sexually mature individual moving between the populations i and j . The model is implemented in *R* using the ‘*glads*’ package, further details of which can be found in Quilodr  n et al. (2020).

4.2.2 Selection on introgressed loci

We simulated two populations, each representing a diverging lineage, between which we measured the level of introgression. Each of these lineages were fixed for an alternate allele; so, by simulating uni-directional migration from the donor to the recipient

population, the proportion of introgression could be calculated by measuring the donor allele frequency in the recipient population. For each lineage, the breeding values for the alleles were drawn from a Gaussian distribution, and by varying the mean of this distribution, we simulated introgression under three scenarios: i) equivalent introgression, ii) positive introgression, and iii) negative introgression. To simulate equivalent introgression, the breeding values for both lineages were drawn using the optimal breeding value ($b_1/2$) as the mean (i.e., under the additive fitness function will form the maximum fitness). Here, the term equivalent introgression is used over neutral introgression, because the populations are still under selection; although, the selection pressure is the same for the two lineages. To simulate positive and negative introgression, the breeding values of the two lineages were drawn from different distributions. One lineage was drawn from a distribution with a mean of $b_1/2$ (the optimal breeding value). The other lineage was drawn from a distribution with a mean of $b_1/2 + \delta_b$, with δ_b defining the deviation from the optimal breeding value. Here, this deviation (δ_b) indicates how much lower the fitness is for the lineage. Thus, for positive introgression, the former distribution was used for the fitter donor population, while the latter was used for the less fit recipient population. The reverse was simulated for negative introgression, where the latter sub-optimal distribution was used for the donor population, and the former optimal distribution was used for the recipient population.

4.2.3 Genes of large effect

To simulate genes of large effect, we weighted the alleles of loci under selection. This was simulated by multiplying the alleles of each selected locus by a relative weight, the sum of which was 1. Genes of large effect were simulated to have a larger than average contribution to the additive phenotype. Therefore, a gene of large effect was any selected

allele where the breeding value was multiplied by a weight greater than the average weight of a selected allele ($1/n_a$). We then varied the frequency of simulated genes of large effect across different parts of the simulated genome (i.e. the number of genes of large effect compared to non-large effect selected loci). Thus, genes of large effect at a high-frequency referred to 30-loci regions that were simulated with 25 genes of large effect and 5 non-large effect genes (Fig. 1). While genes of large effect at a low-frequency referred to 30-loci regions with 5 genes of large effect and 25 non-large effect genes (Fig. 1).

4.2.4 Varying the fitness of hybrids

We included a sex chromosome region with the same size and recombination rate in all individuals. The sex identity of offspring was defined by the sex ratio parameter included in the original *glads* model. However, the sex-specific interactions were based on the inherited copy of the sex chromosome. We introduced the option of modelling sex-specific heterosis and outbreeding depression, where the fitness of a hybrid was respectively higher or lower than the additive fitness. Using a non-recombining sex chromosome region to define if an individual is a hybrid (i.e. when the two sex chromosome haplotypes are from different lineages), we implemented a hybrid-fitness parameter (H) which modified the fitness of hybrids on a sex-specific basis. For each hybrid individual, the product of the fitness (ω) (calculated using eq.2) and the parameter value for that sex gives the modified fitness. For this parameter, a value of 1 denotes a hybrid with a fitness equal to that given by eq. 2 (i.e., the additive fitness), a value of less than 1 denotes a hybrid with a fitness less than that given by eq. 2 (i.e., inbreeding depression), and a value of greater than 1 denotes a hybrid with a fitness greater than that given by eq. 2 (i.e., heterosis).

4.2.5 Initialisation

The simulations implemented in ‘*glads*’ consisted of two diploid populations of 400 individuals with 350 bi-allelic loci, and a 50:50 sex ratio for both lineages. Of the simulated loci, 200 loci contributed to the phenotype additively. Out of the 200 selected loci, 60 loci were simulated as genes of large effect and explained 60% of the additive phenotype, and the remaining genes of regular effect explained 40% of the additive phenotype. These loci were unevenly distributed across 4 regions of 30 loci with varying linkage and frequency of genes of large effect: i) linkage with 25 genes of large effect and 5 genes of regular effect (high frequency), ii) linkage with 5 genes of large effect and 25 genes of regular effect (low frequency), iii) unlinked with 25 genes of large effect and 5 genes of regular effect, and iv) unlinked with 5 genes of large effect and 25 genes of regular effect (Fig. 1). The remaining 150 loci were considered to be neutral. Details of the parameters for the simulations can be found in Table 1.

Three different recombination rates were simulated: i) linked loci (150 loci, $\rho = 0.001$), ii) unlinked loci (150 loci, $\rho = 0.5$), and iii) a non-recombining sex chromosome region (50 loci, $\rho = 0$) (Fig. 1). A uni-directional migration ($m_{ij} = 5 \times 10^{-4}$) rate was used in order to prevent an individual migrating and then migrating back to its population of origin. Mutation was not included in the simulations.

For our simulations of positive and negative introgression, a fitness difference (δ_b) of 0.4 was applied to the breeding values of the two lineages. The breeding values of both lineages were drawn from a Gaussian distribution with a variance (σ_{bv}) of 0.15. To maintain population size of approximately 400 individuals, a maximum fitness (b_0) of 4 and density dependence (b_3) of 0.005 or 0.004 was used (Table 1).

Using a non-recombining sex chromosome and the hybrid-fitness parameter, we simulated three different hybrid scenarios (i.e., when the two sex chromosome haplotypes were from different populations): i) Haldane's rule, where the fitness of hybrid individuals for the heterogametic sex was 40% of the fitness given by eq.2; ii) a full gene flow barrier, where the fitness of hybrid individuals for both sexes was 40% of the fitness given by eq.2; iii) no decrease in hybrid fitness, where the fitness of hybrids for the non-recombining sex chromosome was equal to the fitness given by eq. 2 (Table 1).

4.2.6 Evaluation

Over the course of the simulation, the GLI was recorded in the recipient population. The level of introgression for each locus was calculated as the proportion of donor alleles in the recipient lineage at that locus. A Hidden Markov model (HMM) was used to characterise the GLI. The HMM identified three hidden states: (i) background levels of introgression; (ii) islands of introgression; (iii) valleys of introgression. We simulated eight types of loci: genes of large effect at a high frequency, genes of large effect at a low frequency, genes of regular effect (i.e. non-large effect selected loci), and neutral loci, for linked and unlinked regions. 30 loci were sampled from each of these regions to standardise the HMM, allowing a comparison of landscapes across the different regions. The HMM was simulated using the approach of Marques et al. (2016) in the 'HiddenMarkov' R package (Harte, 2017). For the HMM, a Baum-Welch algorithm was used with 1000 randomly chosen initial parameters to identify the highest likelihood parameter. The Viterbi algorithm then used the highest likelihood parameter to determine the most likely order of states of the landscape of introgression across the simulated loci. Lastly, the significance of the states was calculated by permuting the islands and valleys of introgression 10,000 times across the simulated loci (Quilodrán, Ruegg, et al., 2020b).

4.2.7 Simulation

Our simulations were designed to elucidate how the concurrent interactions of different evolutionary processes influenced the GLI. Our simulations explore the effects of selection, genes of large effect, and hybrid fitness, on the GLI alongside drift and recombination. In this study, we simulated positive, negative, and equivalent introgression, each under i) without a reduced hybrid fitness (i.e. an additive genotype-phenotype map) ii) Haldane’s rule (only the heterogametic sex had a gene flow barrier) and iii) a gene flow barrier for hybrids in both sexes (henceforth referred to as simulations with a full gene flow barrier). All simulations were run for 2000 generations. In the supporting information, we also include the GLI for these simulations after 200 generations (Fig. S1), and the GLI for a simulation which was run with same parameters, but migration was only simulated for the first 200 generations (Fig. S2). Each simulation was independently repeated 100 times, with varying breeding values for each independent simulation (i.e., a different set of breeding values was drawn for each independent simulation from the Gaussian distribution). This was implemented using a different pseudo-random number generator in *R* for each independent simulation (`set.seed` function). These simulations were performed at the University of Geneva on the "Yggdrasil" HPC cluster in *R* (v4.2.0; R Core Team, 2022). The code for simulating and measuring the GLI using the modified “*glads*” functions can be found in the supporting information (Appendix S1).

4.3. Results

4.3.1 Adaptive and maladaptive introgression

We performed simulations with a fixed migration rate ($m_{ij} = 5 \times 10^{-4}$) over 2000 generations, while varying the breeding values of the donor and recipient populations to simulate positive or negative introgression. Irrespective of the simulated fitness of hybrids (additive, Haldane's rule, and a gene flow barrier) and type of introgression i.e., equivalent, positive and negative, the 100 independent simulations for each scenario, showed highly variable levels of introgression for autosomal loci. Similarly, under all both positive and negative introgression, islands and valleys of introgression were observed across all autosomal loci, regardless of the recombination rate. Under equivalent introgression on autosomal loci, introgression levels across the 100 independent simulations ranged between fixation of the donor population allele (introgression proportion of 1) to fixation of the recipient population allele (introgression proportion of 0). Under simulations with positive introgression, all unlinked and linked neutral autosomal loci, like with equivalent introgression, showed introgression levels across the 100 iterations ranging from the fixation of the donor to recipient population allele (Fig. 2a-f). Under negative introgression, fixation of the recipient population allele was seen across all autosomal loci (Fig. 2g-i). Linkage affected the variation in levels of introgression at autosomal loci, in which unlinked loci showed more variation in levels of introgression than linked loci (Fig. 2g-i).

When comparing the average levels of introgression computed for each locus over the 100 independent simulations, selected loci under positive introgression, where the breeding values of the donor population conveyed a fitness advantage over the recipient

population, had the highest level of introgression across the autosomal loci (Fig. 2a-c). However, our simulations of equivalent introgression, where the donor and host population had the same breeding values and fitness, all autosomal loci showed average levels of introgression similar to unlinked neutral loci under positive introgression (Fig. 2a-f). Introgression levels for linked autosomal loci were lowest under negative introgression i.e. when there was a maladaptive effect in the recipient population (Fig. 2).

The frequency of islands and valleys of introgression across the 100 independent simulations differed between selected (i.e., loci that contributed to the additive phenotype and fitness) and neutral loci. The HMM showed that under positive introgression, islands formed more frequently and valleys less frequently than for neutral loci (Fig. 3a-c). Under equivalent introgression, the HMM model showed that islands and valleys were evenly distributed between selected and neutral loci (as was the case for all locus types under equivalent introgression). Under negative introgression, the HMM did not show a clear pattern between selected and neutral loci; although, under Haldane's rule, islands formed more frequently under selected loci.

4.3.2 Fitness of hybrids

Our simulations showed that decreased hybrid fitness (i.e., a full gene flow barrier and Haldane's rule) did not affect the average shape of the GLI but resulted in a decreased level of introgression (Fig. 2). This decrease was larger under simulations of a full gene flow barrier than Haldane's rule (Fig. 2). When analysing the variation across the 100 independent simulations, for all three types of introgression, a lower level of variation was observed for simulations with a reduced hybrid fitness (with simulations under Haldane's rule showing less variation than simulations with a full gene flow barrier) (Fig. 2). For all autosomal loci, under all three types of introgression, a similar trend was also

observed with the average landscape of introgression: that simulations with a reduced hybrid fitness had a lower level introgression (with simulations under Haldane's rule showing lower average levels of introgression than simulations with a full gene flow barrier) (Fig. 2).

4.3.3 Genes of large effect

While the level of introgression was highly variable in each independent simulation, the effect of frequency of genes of large effect on the GLI was influenced by the type of introgression and recombination rate (Fig. 2). In linked regions under positive introgression, both regions of high and low-frequency of genes of large effect, on average, showed either fixation or near fixation of the donor population alleles in the recipient population (Fig. 2a-c). However, more loci in the high-frequency region were shown to fix for the donor allele (Fig. 2a-c). The HMM for linked loci showed that islands and valleys were evenly distributed between the high and low-frequency regions (Fig. 3a-c). In unlinked regions under positive introgression, the average landscape of introgression fluctuated highly. However, the range of these fluctuations was larger under a high-frequency than low-frequency (Fig. 2a-c). This pattern was shown by the HMM, where at unlinked high-frequency regions in comparison to unlinked low-frequency regions, islands formed more often and valleys less often (Fig. 3a-c). Under equal introgression, regions of high and low-frequency showed similar levels of introgression, on average, irrespective of the recombination rate (Fig. 2d-f). This was shown by the HMM, where islands and valleys were equally distributed between the high and low-frequency regions (Fig. 3d-f). Under negative introgression and linkage, on average, the level of introgression was higher for low-frequency than high-frequency regions of genes of large

effect (Fig. 2g-i). However, at unlinked loci under negative introgression, the frequency of genes of large effect did not affect the GLI (Fig. 2g-i).

Under negative introgression for genes of large effect, the GLI fluctuated less under linkage, with higher levels of introgression forming in independent simulations at unlinked loci. Additionally, the average GLI was higher for unlinked genes of large effect than linked genes of large effect (Fig. 2.g-i). The HMM for negative introgression did show differences in the frequency of island and valley formation between high and low-frequency; although, a clear pattern was not observed (Fig. 3g-i).

4.3.4 Recombination rate

Our simulations showed that the effect of recombination on the GLI was dependent on the type of introgression (Fig. 2). Under positive introgression, linked loci had a higher average level of introgression than unlinked loci (Fig. 2a-c). This pattern was shown by the HMMs, where islands formed more frequently and valleys less frequently when loci were linked vs unlinked (Fig. 3a-c). Furthermore, across the 100 independent simulations, linked loci showed less variation. This was particularly the case for genes of large effect, where under a high frequency, the majority of loci were fixed for the donor allele across all 100 independent simulations. Under equivalent introgression, recombination did not influence the average GLI for the simulated autosomal loci (Fig. 2d-f). The HMM reflected this result as islands and valleys were equally distributed between linked and unlinked loci (Fig. 3d-f). However, under negative introgression, the effect of linkage was dependent on the simulated hybrid fitness scenario (Fig. 2d-f). On average, under simulations without a full gene flow barrier and Haldane's rule, unlinked loci had higher levels of introgression than linked loci (Fig. 2d-f). When considering a full gene flow barrier, linkage did not affect the average GLI for autosomal loci (Fig. 2g-i). However, a

clear pattern was not observed with the HMM, although islands formed more frequently under unlinked loci for simulations without a decrease in hybrid fitness (68.7%) and Haldane's rule (74.7%) (Fig. 3g,i). Also, a bias in valley formation was only observed under a full gene flow barrier, where valleys formation was slightly more frequent for unlinked loci (58.7%) (Fig. 3h).

The introgression level for the non-recombining sex chromosome ($p = 0$) did not depended on the type of introgression for simulations with a reduced hybrid fitness (Fig. 2). Under all simulations with a reduced hybridisation fitness, the loci representing a sex chromosome did not show any introgression, across the 100 independent simulations (Fig. 2). Also, for each independent simulation, irrespective of the hybrid fitness and type of introgression, a fixed introgression level was observed for all 50 loci in the non-recombining sex chromosome (Fig. 2). For simulations without a reduced hybrid fitness, the type of introgression did influence the average level of introgression and variation across the 100 independent simulations. Under positive and equivalent introgression without a reduced hybrid fitness, the level of introgression varied highly between the independent simulations (ranging from the fixation of both the donor and host allele), in contrast to negative introgression which showed less variation (ranging from 0 to ~ 0.5) (Fig. 2c,f,i). Also, positive and equivalent introgression showed a similar average level of introgression (0.64 and 0.60), which was higher than negative introgression (0.05) (Fig. 2c,f,i).

4.4. Discussion

Using an individual-based modelling approach, we show that the GLI is influenced by the type of introgression, the fitness of hybrids and recombination rate. Here, our simulations highlight that the GLI is highly variable and may form islands and valleys under the same parameters. Thus, our results support empirical studies that show the effect of recombination on the landscape of introgression is variable and dependent on evolutionary processes such as selection (Duranton & Pool, 2022). Previous studies have highlighted that islands and valleys of introgression are indicative of adaptive and maladaptive introgression, respectively (Schumer et al., 2018; Setter et al., 2020). Our simulations show that under simulations of positive and negative introgression, selected loci have a greater propensity to form islands and valleys of introgression, respectively. Additionally, our simulations provided theoretical support for empirical studies which showed that Haldane's rule reduced introgression between populations (Ottenburghs, 2022). When simulating Haldane's rule under positive introgression, we also show that islands of introgression are more pronounced under the average GLI. Also, our simulations show extensive stochasticity in the GLI (Fig. 2). This can be partially mitigated by specific initialisation for empirical datasets, where the starting populations are defined by sequence data. However, a key aspect of our results is the inherent stochasticity associated with the GLI. Models that consider a complete or close to a complete genotype-phenotype map along with the inclusion of multiple other biological mechanisms may produce far less stochasticity, but integrating such complexity into the *glads* model is not tractable.

4.4.1 Valleys and islands of introgression can form irrespective of the type of introgression

While our simulations showed that varying the type of introgression influences the landscape of introgression, we also observed that across the independent simulations any pattern could form across the GLI. Previous studies have shown that under adaptive and maladaptive introgression, islands and valleys of introgression frequently form, respectively (Schumer et al., 2018; Setter et al., 2020). While this may be the case, here we demonstrated that other evolutionary processes can generate these patterns in the GLI. For example, our simulations of equivalent introgression highlighted the effect of stochasticity on the GLI, where islands and valleys were observed across the GLI. Under equivalent introgression, the introgressed alleles under selection had a fitness equal to the recipient population. Thus, the effect of selection was equivalent for the populations, and the landscape of introgression could be explained by the other simulated evolutionary processes (Svedberg et al., 2021). We explained the variation across the independent simulations as a consequence of stochasticity, supporting previous studies that highlight the effect of drift on generating heterogeneous GLI (S. H. Martin & Jiggins, 2017). Under our simulations of positive introgression, the fitness advantage at selected loci caused selection to maintain the introgressed alleles (Hedrick, 2013), generating islands on the average GLI. On the other hand, the fitness disadvantage generated by negative introgression caused the introgressed alleles to be purged from the population (Veller et al., 2023), generating the reverse valley on the average GLI. We therefore reconcile our results with empirical association between adaptive and maladaptive introgression, and islands and valleys of introgression, by observing the average landscape of introgression at selected loci. However, we show that interpreting a given pattern on the GLI has to be

done with caution, since the complex interplay between different evolutionary processes can influence islands and valley formation.

Despite the variable outputs of the independent simulations in the GLI, our simulations of Haldane's rule and a full gene flow barrier generated more distinct evidence of islands under positive introgression. The islands of introgression that formed were easier to identify as the strength of the hybridisation barrier increased: the genomic islands were more pronounced with a full gene flow barrier than Haldane's rule. We therefore suggest that the introgression barriers were more permeable to loci favoured by selection, while neutral loci were hindered more by the introgression barrier. While this pattern was hidden by the total variation across the 100 independent simulations, the average level of introgression reflected numerous empirical studies that have shown evidence of lower levels of introgression as a consequence of Haldane's rule (for a review see Ottenburghs, 2022). We therefore suggest that adaptive introgression is easier to detect in the presence of hybrids with a gene flow barrier

In previous two-population models, the GLI evolving under adaptive introgression has been shown to resemble a selective sweep (Santiago & Caballero, 2005; Svedberg et al., 2021). In these simulations, the population with the advantageous allele resembled a hard-selective sweep (i.e. fixation of the allele), while the recipient population resembled a soft selective sweep (i.e. both alleles persist at intermediate frequencies) (Setter et al., 2020). Similarly, in our simulations of positive introgression at unlinked loci, we observed a soft selective sweep pattern, as shown by the proportion of introgression ranging from 0 to 1, which indicates that a single allele did not typically fix, but rather intermediate allele frequencies (represented by the observed intermediate proportion of introgression) were typically observed (Fig. 2). However, a strong selective sweep was observed at linked loci. This is shown by the skew in the proportion of

introgression for higher values. Across all the linked loci, an introgression proportion of 1 was observed, while at few loci an introgression proportion of 0 was observed across the independent simulations and the minimum level of introgression at selected loci was typically high, indicating that the loci typically fixed for the introgressed beneficial allele (Fig. 2). We suggest, that the linkage between these positively selected loci cumulatively increased the effect of positive selection on the donor alleles, leading to its fixation. Our simulations therefore provide additional support for selective sweeps under adaptive introgression. Additionally, we highlight that role of recombination and its interaction with evolutionary processes in the formation of these selective sweeps. Through this interaction, we show how both strong and soft sweeps can form within a single population, as a consequence of genomic architecture.

4.4.2 Genes of large effect and the gene flow-selection-drift balance

Our simulations reflected Orr's (1998) hypothesis, that large effect genes can fix when the population is far from the phenotypic optimum. In our simulations of positive introgression, genes of large effect frequently fixed. In these simulations, the host population was far from the phenotypic optima (due to less fit alleles being simulated in the host population), and thus the introgressed alleles helped the individuals move towards the optima. Numerous empirical studies have identified genes of large effect which contribute to the adaptive evolution of species, examples of which include, pollinator shifts in plants (Bradshaw Jr & Schemske, 2003; Hoballah et al., 2007); insect colouration in *Heliconius* butterflies and Sawfly larvae (Linnen et al., 2018; Livraghi et al., 2021); and pigmentation in mice (Hoekstra et al., 2006). We explain the increased fixation of gene of large effect under linkage, by recombination increasing the cumulative effects of selection. Empirically, this pattern has been observed between small-effect loci,

that through reduced recombination in an inversion, generate increased selective effects (Olson-Manning et al., 2012; Twyford & Friedman, 2015). As predicted by the gene flow-selection-drift balance, alleles under stronger selection are likely to have a longer persistence, which is reflected by the increased fixation of linked loci (Yeaman & Whitlock, 2011). This logic can be used to explain the reverse pattern observed under negative introgression: more variation and higher average level of introgression at unlinked loci. Under negative introgression, the host population is close to the phenotypic optima (due to more fit alleles being simulated in the host population). Thus, explaining the low level of introgression at gene of large effect. Similarly, under negative introgression, linked loci are expected to be under stronger negative selection, and so are likely to persist for a shorter period of time. This compounding effect through linkage for genes of large effect, can be tested using the *glads* model by comparing simulations of linked regions with a single large effect gene to a region with multiple large effect genes. This comparison can then be replicated without linkage: simulating unlinked regions with a single large effect gene and with multiple large effect genes. We anticipate the simulations with linkage will recapitulate our results, while the simulations without linkage (single and multiple genes of large effect) will show more similar levels of introgression compared to simulations with linkage. Our simulations of genes of large effect under equivalent introgression, highlighted the role of migration rate and drift on the GLI. Under equal introgression, as selection was equal for both alleles, the gene flow-selection-drift balance reflected the stochastic GLI, and the average level of introgression reflected the level of gene flow.

4.4.3 Reduced introgression in the non-recombining sex chromosomes

Non-recombining heterogametic sex chromosomes typically have a lower introgression rate than recombining autosomes (Veller et al., 2023). Factors causing this include an increased accumulation of incompatibilities and reduced effective population size (S. H. Martin et al., 2019; Van Belleghem et al., 2018). Our simulations of the non-recombining sex chromosome, which included a reduced hybrid fitness, also had a lower level of introgression than the recombining autosomal loci. The occurrence of both sex chromosome haplotypes within an individual reduced the fitness (i.e. an individual that had both the host and donor haplotype), much like how Dobzhansky Muller incompatibilities mutations from different lineages reduce the fitness through negative epistatic interactions. Thus, generating the same effect of a lower introgression rate (Fusco & Uyenoyama, 2011). We also expected this pattern for simulations without a full gene flow barrier, as the absence of recombination in the heterogametic sex is a known factor in reducing introgression (Fraïsse & Sachdeva, 2021). As for the autosomal chromosomes, the interplay of several population genetics factors, including effective population size, migration rate, mutation rate, and the large X-effect, make predicting the introgression rate in a sex chromosome challenging (Van Belleghem et al., 2018). Our simulations with a reduced hybrid fitness reflect this complexity through the variation observed across the 100 independent simulations for the non-recombining sex-chromosome.

4.5. Conclusion

Interpreting the genomic landscape of introgression must be made with caution (S. H. Martin & Jiggins, 2017), as our simulations show that different levels of introgression can occur under the same evolutionary conditions. Nevertheless, the average trends observed in our simulations add to the theoretical evidence and support empirical studies (Setter et al., 2020; Veller et al., 2023), which associate islands of introgression with adaptive introgression and valleys of introgression with maladaptive introgression (Schluter, 2009; Schumer et al., 2018). Furthermore, our results also highlight the importance of considering different layers of complexity when modelling the GLI, as the interaction of different factors, such as selection, drift, and recombination can concurrently influence the shape of the GLI (Duranton & Pool, 2022). Through computational approaches, such as approximate Bayesian computation and machine learning, empirical observations can be compared with theoretical expectations, such as the one provided by our modelling approach, to narrow down the range of conditions that generate particular GLI. While this will not alleviate the inherent stochasticity generated by the *glads* model, summary statistics inferred from the average genomic landscape, using these computational approaches, can still assign probabilities for parameters in forming a particular genomic landscape (Csilléry, 2010). Thus, implementing these methods to our model framework can provide insight into the evolutionary processes shaping the GLI of any biological system.

4.6. References

- Adavoudi, R., & Pilot, M. (2022). Consequences of hybridization in mammals: A systematic review. *Genes*, 13(1). <https://doi.org/10.3390/genes13010050>
- Barton, N., & Bengtsson, B. O. (1986). The barrier to genetic exchange between hybridising populations. *Heredity*, 57(3), 357–376.
- Barton, N. H. (1983). Multilocus clines. *Evolution*, 454–471.
- Bradshaw Jr, H. D., & Schemske, D. W. (2003). Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature*, 426(6963), 176–178.
- Csilléry, K., Blum, M. G. B., Gaggiotti, O. E., & François, O. (2010). Approximate Bayesian Computation (ABC) in practice. In *Trends in Ecology and Evolution* (Vol. 25, Issue 7, pp. 410–418). <https://doi.org/10.1016/j.tree.2010.04.001>
- Dagilis, A. J., & Matute, D. R. (2023). The fitness of an introgressing haplotype changes over the course of divergence and depends on its size and genomic location. *PLoS Biology*, 21(7), e3002185.
- Duranton, M., & Pool, J. E. (2022). Interactions between Natural Selection and Recombination Shape the Genomic Landscape of Introgression. *Molecular Biology and Evolution*, 39(7). <https://doi.org/10.1093/molbev/msac122>
- Edelman, N. B., & Mallet, J. (2021). Prevalence and Adaptive Impact of Introgression. *Annual Review of Genetics*, 55, 265–283. <https://doi.org/10.1146/annurev-genet-021821-020805>
- Fraïsse, C., & Sachdeva, H. (2021). The rates of introgression and barriers to genetic exchange between hybridizing species: Sex chromosomes vs autosomes. *Genetics*, 217(2). <https://doi.org/10.1093/GENETICS/IYAA025>

- Fusco, D., & Uyenoyama, M. K. (2011). Sex-specific incompatibility generates locus-specific rates of introgression between species. *Genetics*, 189(1), 267–288. <https://doi.org/10.1534/genetics.111.130732>
- Haller, B. C., Galloway, J., Kelleher, J., Messer, P. W., & Ralph, P. L. (2019). Tree-sequence recording in SLiM opens new horizons for forward-time simulation of whole genomes. *Molecular Ecology Resources*, 19(2), 552–566. <https://doi.org/10.1111/1755-0998.12968>
- Haller, B. C., & Messer, P. W. (2019). SLiM 3: Forward Genetic Simulations Beyond the Wright-Fisher Model. *Molecular Biology and Evolution*, 36(3), 632–637. <https://doi.org/10.1093/molbev/msy228>
- Harte, D. (2017). *HiddenMarkov: Hidden Markov Models* (R package version 1.8-11). Statistics Research Associates. <http://www.statsresearch.co.nz/dsh/sslib/>
- Hedrick, P. W. (2013). Adaptive introgression in animals: Examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, 22(18), 4606–4618. <https://doi.org/10.1111/mec.12415>
- Hoballah, M. E., Gübitz, T., Stuurman, J., Broger, L., Barone, M., Mandel, T., Dell’Olivo, A., Arnold, M., & Kuhlemeier, C. (2007). Single gene-mediated shift in pollinator attraction in *Petunia*. *Plant Cell*, 19(3), 779–790. <https://doi.org/10.1105/tpc.106.048694>
- Hoekstra, H. E., Hirschmann, R. J., Bunday, R. A., Insel, P. A., & Crossland, J. P. (2006). A Single Amino Acid Mutation Contributes to Adaptive Beach Mouse Color Pattern. *Science*, 313(July), 101–104. <http://www.ncbi.nlm.nih.gov/pubmed/16825572>
- Jones, M. R., Mills, L. S., Alves, P. C., Callahan, C. M., Alves, J. M., Lafferty, D. J. R., Jiggins, F. M., & Jensen, J. D. (2018). in Snowshoe Hares. *Science*, 1358(June), 1355–1358.

- Kim, B. Y., Huber, C. D., & Lohmueller, K. E. (2018). Deleterious variation shapes the genomic landscape of introgression. *PLoS Genetics*, *14*(10), 1–30. <https://doi.org/10.1371/journal.pgen.1007741>
- Linnen, C. R., O’Quin, C. T., Shackleford, T., Sears, C. R., & Lindstedt, C. (2018). Genetic basis of body color and spotting pattern in redheaded pine sawfly larvae (*Neodiprion lecontei*). *Genetics*, *209*(1), 291–305. <https://doi.org/10.1534/genetics.118.300793>
- Livraghi, L., Hanly, J. J., Van Bellghem, S. M., Montejo-Kovacevich, G., van der Heijden, E. S. M., Loh, L. S., Ren, A., Warren, I. A., Lewis, J. J., Concha, C., Hebberecht, L., Wright, C. J., Walker, J. M., Foley, J., Goldberg, Z. H., Arenas-Castro, H., Salazar, C., Perry, M. W., Papa, R., ... Jiggins, C. D. (2021). Cortex cis-regulatory switches establish scale colour identity and pattern diversity in *heliconius*. *ELife*, *10*, 1–31. <https://doi.org/10.7554/eLife.68549>
- Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology and Evolution*, *20*(5), 229–237. <https://doi.org/10.1016/j.tree.2005.02.010>
- Marques, D. A., Lucek, K., Meier, J. I., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2016). Genomics of Rapid Incipient Speciation in Sympatric Threespine Stickleback. *PLoS Genetics*, *12*(2), 1–34. <https://doi.org/10.1371/journal.pgen.1005887>
- Martin, S. H., Davey, J. W., Salazar, C., & Jiggins, C. D. (2019). Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLoS Biology*, *17*(2), 1–28. <https://doi.org/10.1371/journal.pbio.2006288>
- Martin, S. H., & Jiggins, C. D. (2017). Interpreting the genomic landscape of introgression. *Current Opinion in Genetics and Development*, *47*, 69–74. <https://doi.org/10.1016/j.gde.2017.08.007>
- Norris, L. C., Main, B. J., Lee, Y., Collier, T. C., Fofana, A., Cornel, A. J., & Lanzaro, G. C. (2015). Adaptive introgression in an African malaria mosquito coincident with the

- increased usage of insecticide-treated bed nets. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(3), 815–820. <https://doi.org/10.1073/pnas.1418892112>
- Olson-Manning, C. F., Wagner, M. R., & Mitchell-Olds, T. (2012). Adaptive evolution: evaluating empirical support for theoretical predictions. *Nature Reviews Genetics*, *13*(12), 867–877.
- Orr, H. A. (1998). The population genetics of adaptation: The distribution of factors fixed during adaptive evolution. *Evolution*, *52*(4), 935–949. <https://doi.org/10.1111/j.1558-5646.1998.tb01823.x>
- Ottenburghs, J. (2022). Avian Introgression Patterns are Consistent With Haldane’s Rule. *Journal of Heredity*, *113*(4), 363–370. <https://doi.org/10.1093/jhered/esac005>
- Quilodrán, C. S., Ruegg, K., Sendell-Price, A. T., Anderson, E. C., Coulson, T., & Clegg, S. M. (2020). The multiple population genetic and demographic routes to islands of genomic divergence. *Methods in Ecology and Evolution*, *11*(1), 6–21. <https://doi.org/10.1111/2041-210X.13324>
- Ravinet, M., Yoshida, K., Shigenobu, S., Toyoda, A., Fujiyama, A., & Kitano, J. (2018). The genomic landscape at a late stage of stickleback speciation: High genomic divergence interspersed by small localized regions of introgression. In *PLoS Genetics* (Vol. 14, Issue 5). <https://doi.org/10.1371/journal.pgen.1007358>
- Santiago, E., & Caballero, A. (2005). Variation after a selective sweep in a subdivided population. *Genetics*, *169*(1), 475–483. <https://doi.org/10.1534/genetics.104.032813>
- Schluter, D. (2009). Evidence for ecological speciation and its alternative. *Science*, *323*(5915), 737–741. <https://doi.org/10.1126/science.1160006>

- Schumer, M., Xu, C., Powell, D. L., Durvasula, A., Skov, L., Holland, C., Blazier, J. C., & Sankararaman, S. (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science*, 360(6389), 656–660.
- Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends in Ecology and Evolution*, 19(4), 198–207. <https://doi.org/10.1016/j.tree.2004.01.003>
- Setter, D., Mousset, S., Cheng, X., Nielsen, R., DeGiorgio, M., & Hermisson, J. (2020). VolcanoFinder: Genomic scans for adaptive introgression. In *PLoS Genetics* (Vol. 16, Issue 6). <https://doi.org/10.1371/journal.pgen.1008867>
- Svedberg, J., Shchur, V., Reinman, S., Nielsen, R., & Corbett-Detig, R. (2021). Inferring Adaptive Introgression Using Hidden Markov Models. *Molecular Biology and Evolution*, 38(5), 2152–2165. <https://doi.org/10.1093/molbev/msab014>
- Taylor, S. A., & Larson, E. L. (2019). Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature Ecology and Evolution*, 3(2), 170–177. <https://doi.org/10.1038/s41559-018-0777-y>
- Team, R. C. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.r-project.org/>
- Twyford, A. D., & Friedman, J. (2015). Adaptive divergence in the monkey flower *Mimulus guttatus* is maintained by a chromosomal inversion. *Evolution*, 69(6), 1476–1486. <https://doi.org/10.1111/evo.12663>
- Van Belleghem, S. M., Baquero, M., Papa, R., Salazar, C., McMillan, W. O., Counterman, B. A., Jiggins, C. D., & Martin, S. H. (2018). Patterns of Z chromosome divergence among *Heliconius* species highlight the importance of historical demography. *Molecular Ecology*, 27(19), 3852–3872. <https://doi.org/10.1111/mec.14560>

- Veller, C., Edelman, N. B., Muralidhar, P., & Nowak, M. A. (2023). Recombination and selection against introgressed DNA. *Evolution; International Journal of Organic Evolution*, 77(4), 1131–1144. <https://doi.org/10.1093/evolut/qpad021>
- Yeaman, S., & Whitlock, M. C. (2011). The genetic architecture of adaptation under migration-selection balance. *Evolution*, 65(7), 1897–1911. <https://doi.org/10.1111/j.1558-5646.2011.01269.x>

4.7. Figures

Table 1 – **List of parameters**

Parameter	Definition	Value
N	Population size	400
n _a	Number of loci	350
	Number of selected loci	200
	Number of candidate loci	60 (60% of phenotype)
	Number of non-candidate loci	140 (40% of phenotype)
	Weighting for selected loci	0.01 (genes of large effect) ~ 0.00286 (genes of regular effect) †
δ _b	Selection on introgressed allele	0.4
m _{ij}	Uni-directional migration rate	5 x 10 ⁻⁴ (population 1 to 2)
t	Number of generations	2000
b ₀	Maximum offspring number	4
b ₁	Phenotypic optima	0.4

b_2	Variance of fitness curve	0.8
σ_{bv}	Variance of breeding values	0.15
b_3	Density dependence (host)	0.005 ‡ 0.004 §
σ_{env}	Stochastic environmental variance (host)	0.01
σ_{dem}	Stochastic demographic variance (host)	1
μ	Mutation rate	0
	Sex ratio	0.5
ρ	Recombination rate	0.001 (Linked loci) 0.5 (Unlinked loci) 0 (non-recombining sex chromosome)
H	Haldane's rule	0.4
	A full gene flow barrier	0.4

† 0.4 / 140

‡ for populations where the breeding values were drawn from a Gaussian distribution with a

mean of $b_1/2$

§ for populations where the breeding values were drawn from a Gaussian distribution with

mean of $b_1/2 + \delta_b$

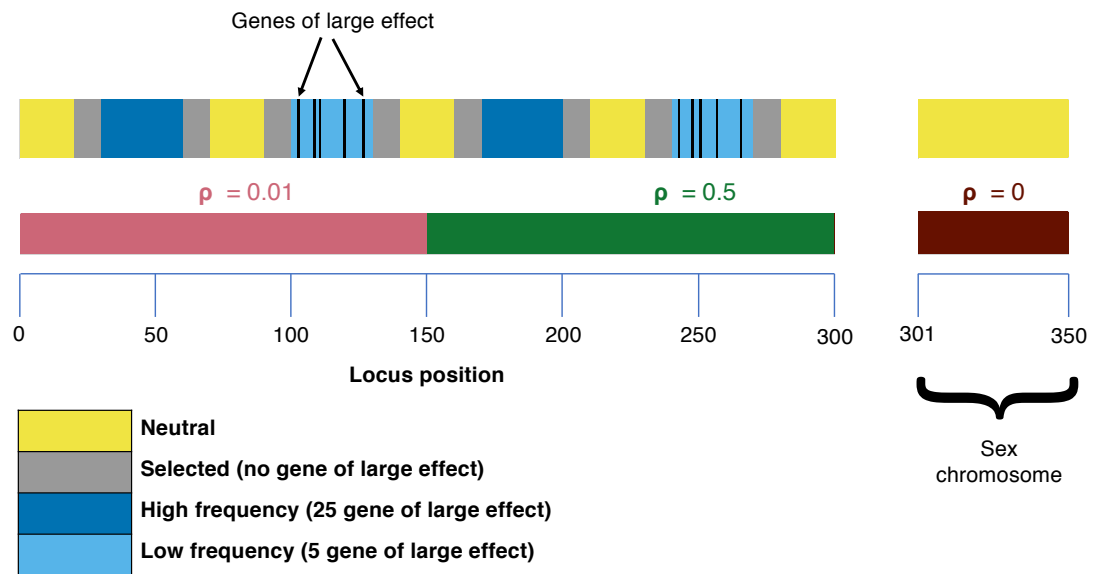


Figure 1. The genetic architecture of the 350 simulated loci. The upper bar describes the type of selection acting on loci. The black horizontal stripes in the low-frequency region represent the position of the 5 candidate loci in this region. The high and low-frequency regions are 30 loci long. The lower bar represents the recombination rate (ρ).

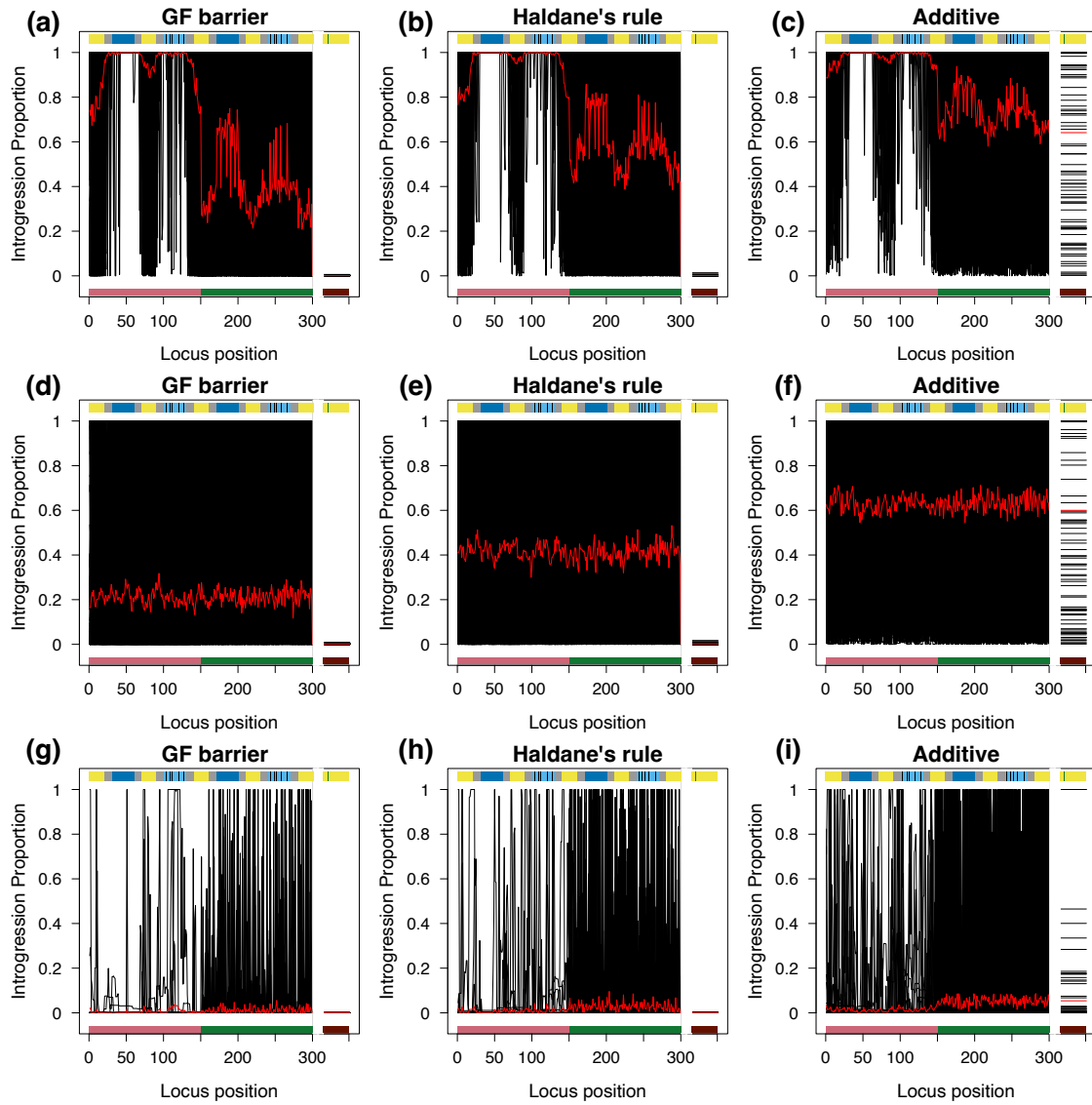


Figure 2. The genomic landscape of introgression (GLI) between two lineages. Each panel represents the variation across 100 independent simulations of 2000 generations, each GLI is represented with black lines. The line in red is the average of the 100 independent simulations. The upper bar represents the selection the locus is under and the lower bar the recombination rate (ρ) (see Figure 1). The different panels show simulations under different types of introgression and hybrid fitness. The top row represents simulations under positive introgression, the middle panel under neutral introgression, and the bottom panel under negative introgression.

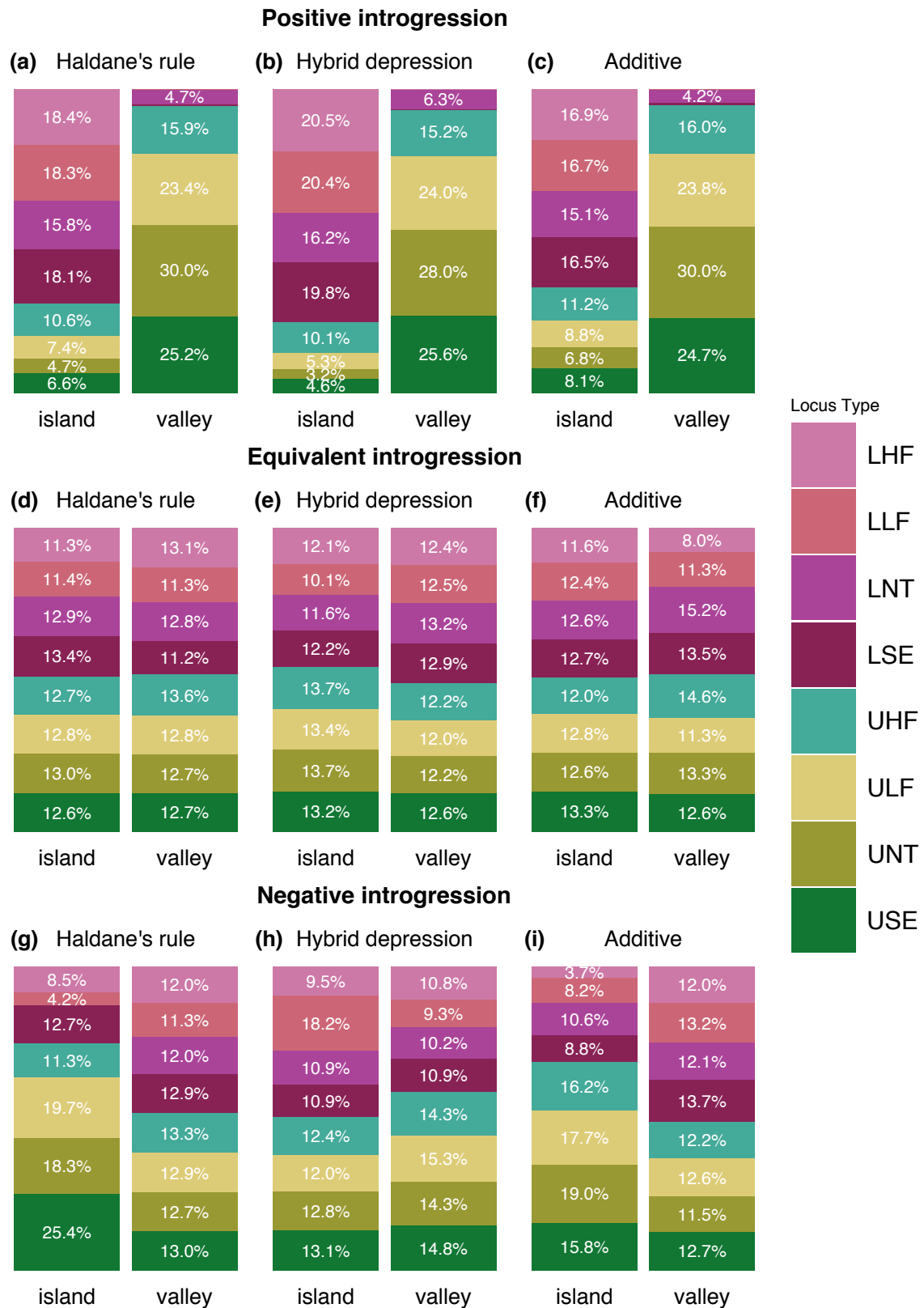


Figure 3. The proportion of islands and valley formation as identified by the hidden Markov model. The different panels show simulations under different types of

introgression and hybrid fitness. Each panel represents the proportion of islands and valleys identified by the HMM for each simulation under the following locus types: linked high frequency (LHF), linked low frequency (LLF), linked neutral (LNT), linked selected (no genes of large effect) (LSE), unlinked high frequency (UHF), unlinked low frequency (ULF) unlinked neutral (UNT), unlinked selected (no genes of large effect) (USE). For values that are $< 1\%$, see Table S1.

4.8 Supporting Information

Table S1. Percentage of islands and valleys identified by the HMM for each simulation after 2000 generations

Landscape		Simulation								
	Locus Type	A.I. + H.Rule	A.I. + Full	A.I. + Add.	N.I. + H.Rule	N.I. + Full	N.I. + Add.	M.I. + H.Rule	M.I. + Full	M.I. + Add.
valley	LHF	0.08	0.18	0.00	13.09	12.38	8.04	11.96	10.79	11.99
	LLF	0.19	0.00	0.28	11.28	12.53	11.33	11.29	9.29	13.16
	LSE	0.51	0.63	0.92	11.18	12.87	13.53	12.86	10.93	13.69
	LNT	4.71	6.27	4.23	12.75	13.24	15.17	12.03	10.25	12.14
	UHF	15.91	15.23	16.02	13.60	12.22	14.63	13.30	14.34	12.22
	ULF	23.42	24.01	23.81	12.75	11.98	11.33	12.86	15.30	12.59
	USE	25.21	25.63	24.73	12.70	12.59	12.61	13.00	14.75	12.71
	UNT	29.96	28.05	30.00	12.65	12.19	13.35	12.71	14.34	11.50
island	LHF	18.42	20.46	16.85	11.26	12.06	11.58	8.45	9.49	3.72
	LLF	18.33	20.36	16.67	11.35	10.07	12.44	4.23	18.25	8.19
	LSE	18.13	19.75	16.55	13.40	12.19	12.67	12.68	10.95	8.75
	LNT	15.81	16.18	15.07	12.87	11.64	12.58	0.00	10.95	10.61
	UHF	10.59	10.10	11.15	12.75	13.74	11.99	11.27	12.41	16.20
	ULF	7.44	5.34	8.81	12.79	13.39	12.76	19.72	12.04	17.69
	USE	6.57	4.63	8.14	12.61	13.19	13.35	25.35	13.14	15.83
	UNT	4.72	3.18	6.75	12.97	13.71	12.62	18.31	12.77	18.99

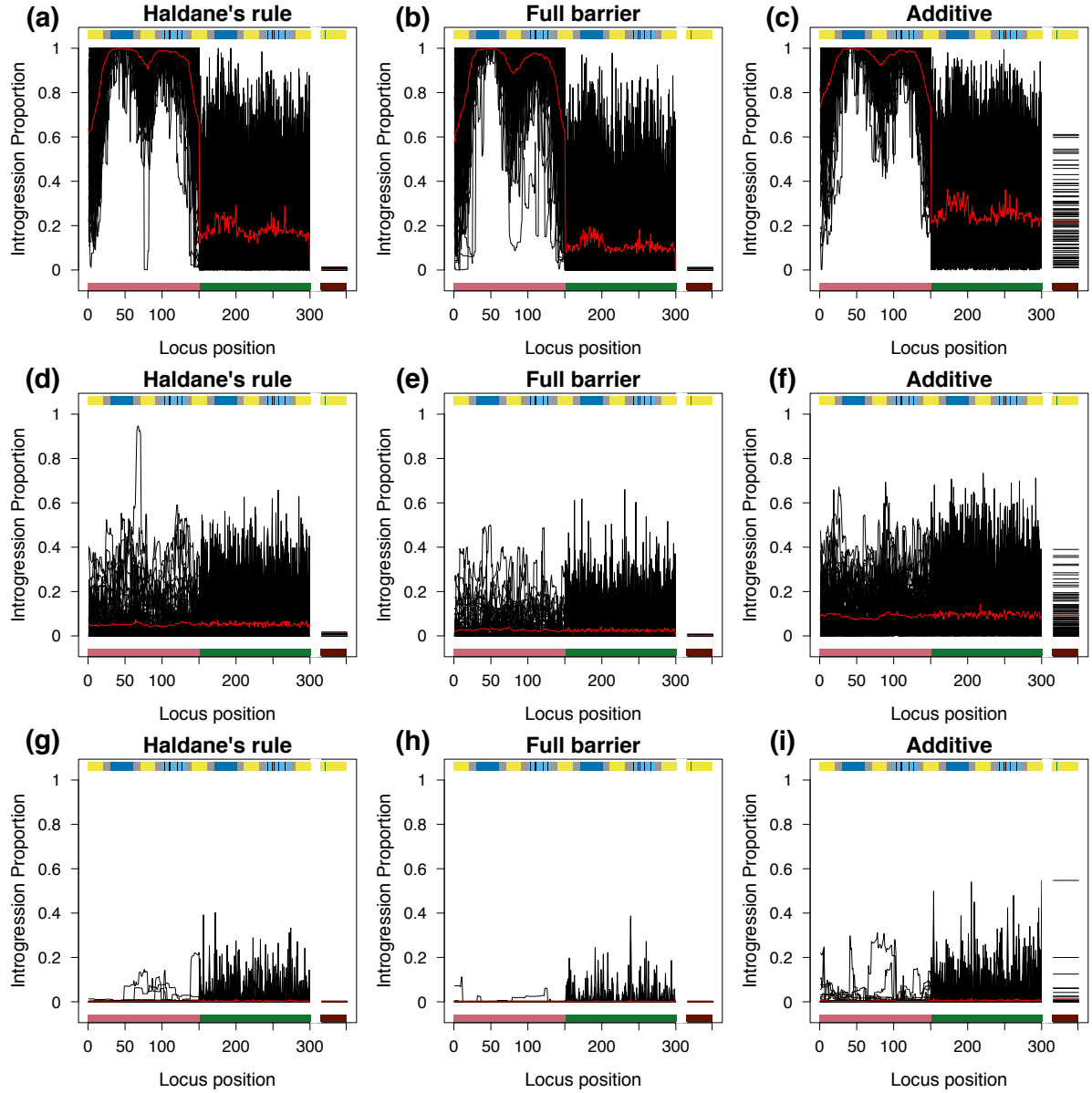


Figure S1. The genomic landscape of introgression (GLI) between two lineages. Each panel represents the variation across 100 independent simulations of 2000 generations after 200 generations. Each GLI is represented with a black lines. The line in red is the average of the 100 independent simulations. The upper bar represents the selection the

locus is under and the lower bar the recombination rate (ρ) (see Figure 1). The different panels show simulations under different types of introgression and hybrid fitness. The top row represents simulations under positive introgression, the middle panel under neutral introgression, and the bottom panel under negative introgression.

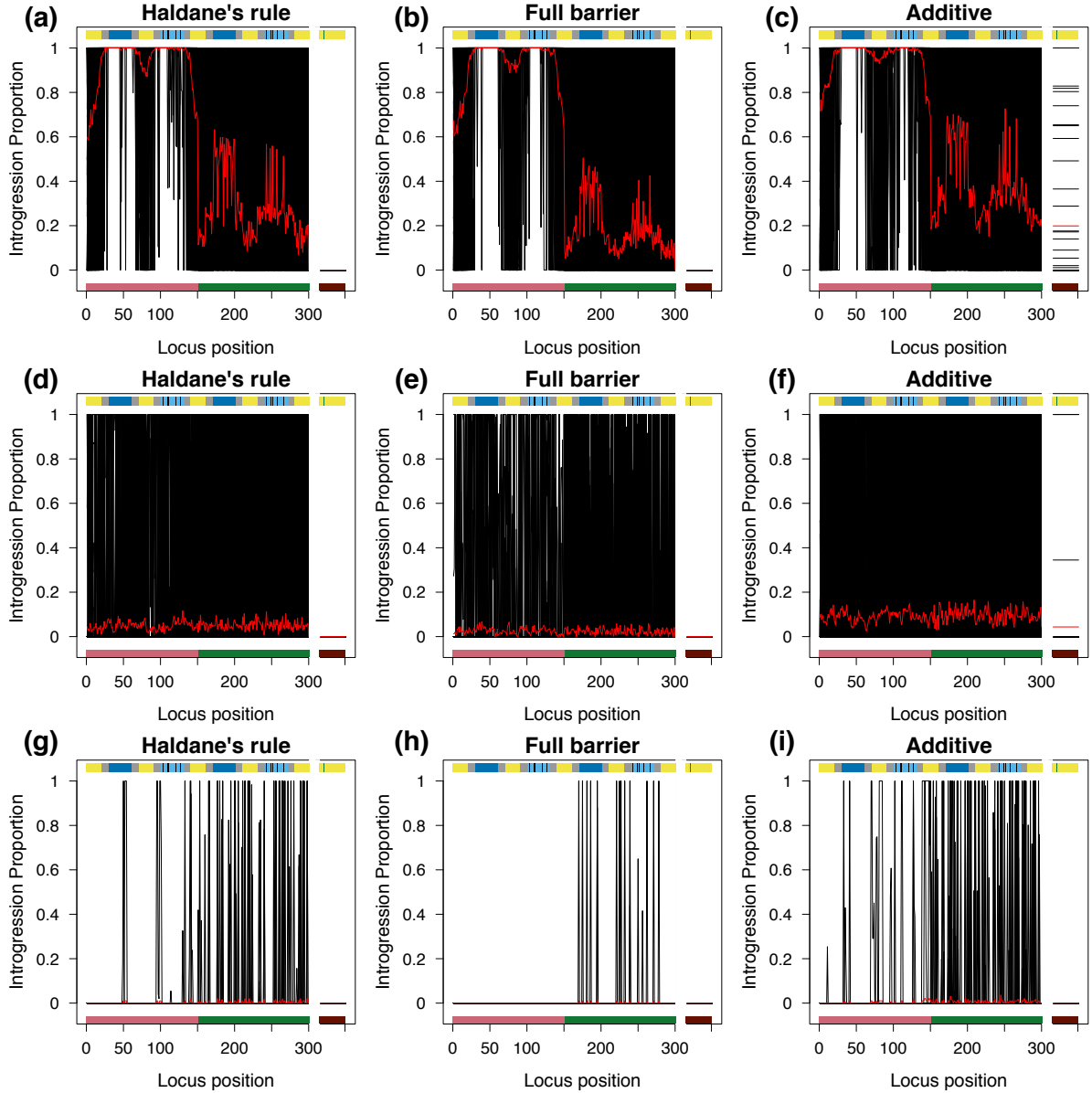


Figure S2. The genomic landscape of introgression (GLI) between two lineages. Each panel represents the variation across 100 independent simulations after 2000 generations (with migration for only the first 200 generations). Each GLI is represented with a black lines. The line in red is the average of the 100 independent simulations. The upper bar

represents the selection the locus is under and the lower bar the recombination rate (ρ) (see Figure 1). The different panels show simulations under different types of introgression and hybrid fitness. The top row represents simulations under positive introgression, the middle panel under neutral introgression, and the bottom panel under negative introgression.

5. Discussion

5.1 General discussion

At the heart of evolutionary biology is the question of how populations diverge and species form. This question inspired Darwin and Wallace to present their historic developments on the theory of natural selection (Darwin & Wallace, 1858). Over 160 years of development in the field has led to an extended understanding of evolution, but it is still challenging to decipher the mechanism participating in the divergence of a particular species. The effects the evolutionary processes selection, mutation, drift, and gene flow must be considered not only individually, but also interactively, as different processes can have antagonistic or synergistic effects depending on the context. In this thesis, I present a series of interlinked models that expand on our ability to dissect the evolutionary histories of species via examination of the genomic landscape. I highlight cautions in interpreting a given genomic pattern as the different evolutionary processes can generate similar genomic landscapes. I also develop an individual-based modelling approach useful for inferring the likely evolutionary processes in play for a specific biological system.

5.2 An improved general model for simulating genomic landscapes

In this thesis, I make use of the individual based *Genomic Landscape of Divergence Simulations* (*'glads'*) model developed by Quilodr  n et al., 2020. This approach enables the simulation of the divergence of two or more genetically defined populations (lineages)

across three hierarchical levels: the genotype, phenotypic traits, and demographic rates. The model proceeds in discrete steps representing generations. In each step, the populations breed together and migrate between the populations. Over the course of the simulation, the aforementioned evolutionary processes: selection, mutation, recombination, drift, and gene flow, shape the evolutionary divergence of the populations. While this thesis focusses on the fixation index (F_{ST}) and level of introgression to understand the impact of evolutionary processes on diverging lineages, any population genomics statistics could be implemented to the output of this general approach. F_{ST} can be inflated by a reduction in within-population genetic diversity (Charlesworth, 1998,). Nevertheless, F_{ST} is particularly informative for inferring the process that generate divergence between populations and maintain/erode this divergence.

I have incorporated important new features to this modelling framework, including: i) an extension of the genotype-phenotype map, with the inclusion dominance, genotype-environment interaction, and epistatic interactions ii) positive and negative frequency dependent selection, iii) variable patterns of migration, in contrast to the fixed migration rate parameter, iv) the simulation of multiple traits within a population, v) a framework for examining the genomic landscape of introgression, and vi) reduced gene flow in hybrids. The implementation of different layers of complexity, simulated with the concurrent influence of these new features, enables a more nuanced characterization of the conditions observed in natural systems.

5.3 Individual based modelling

Individual based models provide an insight at the higher level through the emergent properties of the interactions of parametrised individuals at a lower level. As a result,

analytical solutions for these models are too complex. Thus, inferring the meaning of these emergent properties can be challenging. In this thesis, the formation of genomic landscapes is not explanatory of the biological underpinnings. Rather, comparative approaches between different parameter conditions highlight the effects these parameters can have on the genomic landscape e.g. introducing negative frequency dependent selection transforms genomic islands into valleys. However, the biological underpinnings can be potentially inferred with confidence by looking at the emergent population characteristics e.g. the phenotypic space, and changes in population size. Additionally, such inferences can be influenced by verbal and conceptual understandings of the mechanism e.g. frequency dependent selection has been shown to generate balancing selection empirically with a conceptual basis. Nevertheless, individual based models provide a fruitful insight into higher level dynamics that differential equation-based approaches cannot provide (Breckling, 2002). While differential-based approaches offer mathematical transparency that facilitates parameter generalisations and explicit solutions, they are unable to capture the complexity of heterogeneous system with individuals in the population optimizing their fitness over time (Grimm, 2019). Through the *glads* model, we can simulate the genomic landscape under different evolutionary conditions. These predictions can then be compared to empirical datasets to provide additional lines for already existing hypotheses, by applying the corresponding parameters and checking for congruence between the two (Martin et al., 2024; Appendix 2). However, the reverse approach of using individual-based models to derive an evolutionary narrative only from the empirical genomic landscape is challenging. Individual-based models whilst versatile, do not provide generalised rules, regarding parameter, as found with analytical based approaches e.g. differential equation models. Approaches such as ABC and machine learning can mitigate this to a degree but still do

not provide precise rules regarding the parameter for an observed genomic landscape. Rather, such approaches can narrow the range of conditions that might explain the observed genomic landscapes, through the summary statistics for the observed F_{ST} and introgression landscape distributions. With a more restricted parameter space for a given landscape, certain hypotheses about the evolutionary history of a species can potentially be ruled out. This model can be used to simulate parts of a chromosome using the *geno2glads* tool (see Supplementary material, Appendix 1). Simulating genomic data on a larger scale, while possible, will typically exceed computational restraints, restricting the scale for which simulations are developed.

5.4 The variability of the genomic landscape

Over the last few decades, there has been an interest in studying the divergence of populations from the genomic landscape perspective (Charlesworth et al., 1997; Nosil, Funk, et al., 2009; Wu & Ting, 2004), evaluating regions of elevated and reduced differentiation or similarities across genomes. Early approaches assumed that genomic patterns (i.e., high or low levels of differentiation or introgression) formed by a particular set of evolutionary processes, which could then be used to infer the evolutionary history of a species (Wu, 2001; Wu & Ting, 2004). However, more recently it was observed that the same pattern can form under varying evolutionary conditions (Ravinet et al., 2017). In this thesis, by further extending the complexity of the *glads* model, I aimed to see if it was possible to identifying the range of parameters and processes under which landscape features form. I demonstrated that while different evolutionary processes can generate similar genomic patterns, a single set of evolutionary conditions can also generate multiple different genomic patterns, as a consequence of stochasticity (e.g. genetic drift

and random demographic fluctuations) and genomic composition of the starting population. I thus emphasise caution in interpreting the genomic landscapes of both divergence and introgression. Nevertheless, while the genomic landscape is highly variable, different evolutionary processes and combinations of processes often have higher propensities to form particular genomic patterns e.g. for the landscape of divergence, divergent selection combined with positive frequency-dependent selection has a high propensity for islands formation and a low propensity for valleys formation (Fig. 1).

5.4.1 Genomic landscape of divergence

The modern evolutionary synthesis provides a strong understanding of how different evolutionary processes contribute to divergence along the speciation continuum (Coyne & Orr, 2004a). However, the question remains how do these processes influence divergence at the level of genomes, and can we disentangle the evolutionary history of a specific biological system using patterns on the genomic landscape of divergence. In Chapter 2, I introduce several modifications to the *glads* model, and highlight: 1) the variable nature of the landscape of divergence, and 2) the effects of these additional layers of complexity on the propensity of islands and valleys. Previously, genomic islands were thought to be restricted to the condition of divergent selection and gene flow, where divergent selection generates divergence at selected loci, while gene flow limits divergence at neutral loci through homogenisation (Nosil, Funk, et al., 2009). However, our simulations show that irrespective of divergent or parallel selection between populations, there is a propensity for the formation of genomic islands and valleys of divergence, when negative and positive frequency-dependent selection act within the

populations, respectively. Under the concurrent action of dominance and divergent selection, and genotype-environment interaction with parallel selection, our simulations also show a propensity for reduced levels of divergence. By modelling the effects of concurrent evolutionary processes, our results build on previous work that show how multiple factors, and their concurrent effects, shape the genomic landscape of divergence (Campbell et al., 2018; Quilodr  n, Ruegg, et al., 2020b; Ravinet et al., 2017). By highlighting new routes for the formation of genomic islands and valleys e.g. frequency-dependent selection, our results build on previous research which indicate that genomic islands of divergence can form via different evolutionary routes (Quilodr  n, Ruegg, et al., 2020b; Yeaman et al., 2016).

Dobzhansky-Muller incompatibilities (DMIs) describe the negative epistatic interactions that can form between alleles that arise in different evolutionary backgrounds. Consequently, these DMIs can facilitate divergence and even speciation, through post-zygotic effects on fitness. While the effects of DMIs on reproductive isolation have been modelled, the effects of DMIs on the genomic landscape of divergence had not previously been explored. In Chapter 3, I developed the *glads* mutation rate parameter to incorporate DMIs into the model. Our results showed that with the inclusion of DMIs under simulations with high levels of gene flow, genomic islands had a higher propensity to form. Thus, our results provide support for i) previous theoretical studies which show that DMIs decrease gene flow between populations (Blanckaert & Hermisson, 2018), and ii) the effect of selection on the genomic landscape is counterbalanced, when both divergent and parallel selection concurrently acts on populations. The genomic architecture of species contains multiple genes, which may be differentially affected by selection i.e. some may be under divergent selection and others parallel selection (Langerhans & Riesch, 2013). As shown in Chapter 2, and in other

studies (Quilodr  n, Ruegg, et al., 2020b), selection impacts the genomic landscape of divergence in different ways. While selection has been modelled in terms of the genomic landscape, previous studies have focused on a single trait (Quilodr  n, Ruegg, et al., 2020b), and often simulated selection by using a single parameter (Charlesworth et al., 1997; Feder, Gejji, et al., 2012b; Feder & Nosil, 2009; Sedghifar et al., 2016). In Chapter 3, I showed that the concurrent simulation of two traits under selection influences their effect on genomic landscape of divergence, as a consequence of their interaction.

The discrepancies in the genomic landscape of divergence observed in Chapter 2 and Chapter 3, can be explained by the construction of the model. Across these two chapters, multiple different genotype-phenotype maps and models of selection (i.e. a single trait under natural selection, two traits under natural selection, parallel and divergent natural selection, and positive and negative frequency dependent selection) were simulated. Fundamentally, these different modifications of the model impacted the genomic landscape in different ways. In addition to this, the concurrent effects of these modifications may differ from the effects of these modifications occurring on their own. This is particularly highlighted by the difference in genomic landscape observed in Chapter 2 and Chapter 3, where the addition of second trait significantly changes the effect of selection on the genomic landscape. Furthermore, I anticipate that this effect will occur repeatedly for the modifications introduced in this thesis. Thus, one of the next steps will be integrate the modifications in this thesis, in order to be able to simulate them concurrently. While this additional complexity makes parameter generalisation more difficult and reduces mathematical transparency, the model will represent a more accurate approximation of empirical systems, when parametrised correctly. Additionally, a rich insight can be obtained into the concurrent vs independent effects of these modifications on the genomic landscape of divergence.

The findings of Chapter 2 and 3 together, through simulating an increased complexity of the genotype-phenotype map and fitness function i.e. the addition of frequency-dependent selection and simulation of multiple traits, show novel routes for the formation of particular genomic landscapes of divergence. I thus highlight that implementing additional layers of complexity has a profound impact on the shape of the genomic landscape of divergence. Our independent simulations with equal parameters emphasize the highly variable output on the divergence levels, although certain genomic patterns are more likely to be expected under given conditions (Fig. 1).

5.4.2 Genomic landscape of Introgression

Hybridisation affects many species, with a dynamic range of effect on their evolution, ranging from species fusion to facilitating speciation (Seehausen, 2004). Similar to the genomic landscape of divergence, there is a poor understanding of how different evolutionary processes shape the genomic landscape of introgression, and to what extent it is possible to disentangle the effects of these processes on the evolution of species. In Chapter 4, I developed the *glads* model to concurrently simulate different evolutionary processes, and examined their effects on the genomic landscape of introgression. In this chapter, I provide support for studies that show the effect of selection and linkage on the landscape of introgression (Duranton & Pool, 2022; Ravinet et al., 2018), in which the type of selection influences the effect recombination has on the genomic landscape of introgression. Additionally, our results add to the literature that highlights the propensity of positive and negative introgression to form islands and valleys of introgression, respectively (Schumer et al., 2018; Setter et al., 2020). Here, I show how the concurrent interactions of different evolutionary processes and the genetic architecture (i.e. genes of

large effect and recombination) lead to a highly variable genomic landscape of introgression. Additionally, I show with barriers to gene flow in hybrids, that the average trend of genomic islands of introgression under positive selection becomes more pronounced, thus indicating that such barriers may help to recognise adaptive introgression in empirical studies.

5.5 Conclusion

In each chapter, I examine the effects of different evolutionary processes on the genomic landscapes (divergence and introgression). Across the simulations, I demonstrate that both the genomic landscape of divergence and introgression are highly variable, with different genomic patterns observed under same set of evolutionary conditions. I thus emphasise caution in interpreting genomic patterns. Also, I demonstrate that the inclusion of additional layers of complexity to the *glads* model has a profound effect on the shape of the genomic landscape of divergence and introgression (Nosil, Funk, et al., 2009; Ottenburghs et al., 2017) (Fig. 1). Through computational approaches, such as approximate Bayesian computation and machine learning, the most likely parameter values describing empirical observation could be identified. By combining these tools with the modelling framework presented in this thesis and empirical data, as illustrated in Chapter 3 for the silvereyes (*Zosterops lateralis*), the range of conditions participating in particular biological systems can be narrowed. While the *glads* model shows considerable stochasticity in the genomic landscape, as highlighted in this thesis, the average genomic landscape provides an insight into the effects of different evolutionary mechanisms on the genomic landscape. Machine learning and Bayesian approaches can utilise summary statistics for these average landscapes, producing likelihoods for different parameter

combinations. These likelihoods can be used to infer the most probable parameter spaces for the observed empirical genomic landscapes; thus, alleviating the inherent challenges with interpreting *glads* simulations that generate a large amount of stochasticity. I emphasise that the genomic landscapes of divergence and introgression should concurrently be analysed with multiple interactions between evolutionary processes and layers of complexity, which could dramatically influence any expected genomic pattern. My thesis presents a general framework to address this major questions in the field of evolutionary biology, which may be used to shed light on the past or future evolution of any diverging lineages.

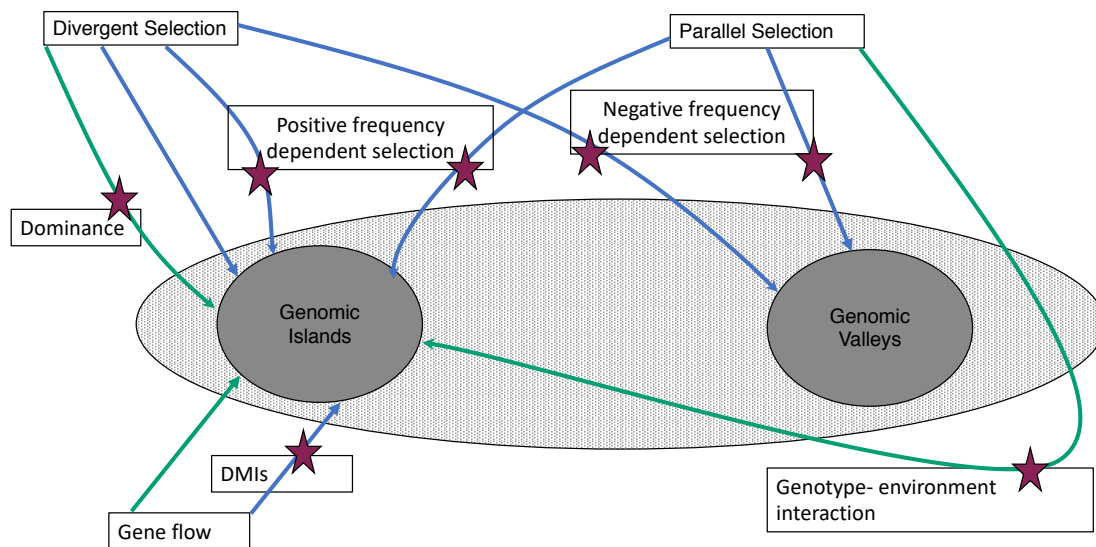


Figure 1. A schematic of the modifications introduced to the model and their effects on the genomic landscape (grey oval rings). The arrows to either the genomic island of divergence or valley of similarity, represents the effect the evolutionary processes on propensity of the landscape to form. The arrows in blue represent an increased propensity for the formation of that landscape, while the arrows in green a reduced propensity.

Arrows going through multiple boxes indicate an interaction (indicated by a purple star). Additionally, genetic drift and the genetic architecture also interact with each of these arrows to influence the genomic landscape of divergence.

5.5 References

- Blanckaert, A., & Hermisson, J. (2018). The limits to parapatric speciation II: Strengthening a preexisting genetic barrier to gene flow in parapatry. *Genetics*, 209(1), 241–254. <https://doi.org/10.1534/genetics.117.300652>
- Breckling, B. (2002). Individual-based modelling: potentials and limitations. In *TheScientificWorldJournal* (Vol. 2, pp. 1044–1062). <https://doi.org/10.1100/tsw.2002.179>
- Campbell, C. R., Poelstra, J. W., & Yoder, A. D. (2018). What is Speciation Genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species. *Biological Journal of the Linnean Society*, 124(4), 561–583. <https://doi.org/10.1093/biolinnean/bly063>
- Charlesworth, B. (1998). Measures of Divergence Between Populations and the Effect of Forces that Reduce Variability. *Mol. Biol. Evol*, 15(5), 538–543. <https://academic.oup.com/mbe/article/15/5/538/987854>
- 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. <https://doi.org/10.1017/S0016672397002954>
- Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sinauer Associates, Inc. Publishers.

- Darwin, C., & Wallace, A. (1858). On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. *Journal of the Proceedings of the Linnean Society of London. Zoology*, 3(9), 45–62.
- Duranton, M., & Pool, J. E. (2022). Interactions between Natural Selection and Recombination Shape the Genomic Landscape of Introgression. *Molecular Biology and Evolution*, 39(7). <https://doi.org/10.1093/molbev/msac122>
- Feder, J. L., Gejji, R., Yeaman, S., & Nosil, P. (2012). Establishment of new mutations under divergence and genome hitchhiking. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1587), 461–474. <https://doi.org/10.1098/rstb.2011.0256>
- Feder, J. L., & Nosil, P. (2009). Chromosomal inversions and species differences: When are genes affecting adaptive divergence and reproductive isolation expected to reside within inversions? *Evolution*, 63(12), 3061–3075. <https://doi.org/10.1111/j.1558-5646.2009.00786.x>
- Grimm, V. (2019). Ecological Models: Individual-Based Models☆. In B. Fath (Ed.), *Encyclopedia of Ecology (Second Edition)* (pp. 65–73). Elsevier. <https://doi.org/https://doi.org/10.1016/B978-0-12-409548-9.11144-3>
- Langerhans, R. B., & Riesch, R. (2013). Speciation by selection: A framework for understanding ecology's role in speciation. *Current Zoology*, 59(1), 31–52. <https://doi.org/10.1093/czoolo/59.1.31>
- Martin, C. A., Sheppard, E. C., Ali, H. A. A., Illera, J. C., Suh, A., Spurgin, L. G., & Richardson, D. S. (2024). Genomic landscapes of divergence among island bird populations: Evidence of parallel adaptation but at different loci? *Molecular Ecology*, 33(12). <https://doi.org/10.1111/mec.17365>

- Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, 18(3), 375–402. <https://doi.org/10.1111/j.1365-294X.2008.03946.x>
- Ottenburghs, J., Megens, H. J., Kraus, R. H. S., Van Hooft, P., Van Wieren, S. E., Crooijmans, R. P. M. A., Ydenberg, R. C., Groenen, M. A. M., & Prins, H. H. T. (2017). A history of hybrids? Genomic patterns of introgression in the True Geese. *BMC Evolutionary Biology*, 17(1), 1–14. <https://doi.org/10.1186/s12862-017-1048-2>
- Quilodrán, C. S., Ruegg, K., Sendell-Price, A. T., Anderson, E. C., Coulson, T., & Clegg, S. M. (2020). The multiple population genetic and demographic routes to islands of genomic divergence. *Methods in Ecology and Evolution*, 11(1), 6–21. <https://doi.org/10.1111/2041-210X.13324>
- Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., Noor, M. A. F., Mehlig, B., & Westram, A. M. (2017). Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, 30(8), 1450–1477. <https://doi.org/10.1111/jeb.13047>
- Ravinet, M., Yoshida, K., Shigenobu, S., Toyoda, A., Fujiyama, A., & Kitano, J. (2018). The genomic landscape at a late stage of stickleback speciation: High genomic divergence interspersed by small localized regions of introgression. In *PLoS Genetics* (Vol. 14, Issue 5). <https://doi.org/10.1371/journal.pgen.1007358>
- Schumer, M., Xu, C., Powell, D. L., Durvasula, A., Skov, L., Holland, C., Blazier, J. C., & Sankararaman, S. (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science*, 360(6389), 656–660.
- Sedghifar, A., Brandvain, Y., & Ralph, P. (2016). Beyond clines: lineages and haplotype blocks in hybrid zones. *Molecular Ecology*, 25(11), 2559–2576. <https://doi.org/10.1111/mec.13677>

- Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends in Ecology and Evolution*, 19(4), 198–207. <https://doi.org/10.1016/j.tree.2004.01.003>
- Setter, D., Mousset, S., Cheng, X., Nielsen, R., DeGiorgio, M., & Hermisson, J. (2020). VolcanoFinder: Genomic scans for adaptive introgression. In *PLoS Genetics* (Vol. 16, Issue 6). <https://doi.org/10.1371/journal.pgen.1008867>
- Wu, C. I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*, 14(6), 851–865. <https://doi.org/10.1046/j.1420-9101.2001.00335.x>
- Wu, C. I., & Ting, C. T. (2004). Genes and speciation. *Nature Reviews Genetics*, 5(2), 114–122. <https://doi.org/10.1038/nrg1269>
- Yeaman, S., Aeschbacher, S., & Bürger, R. (2016). The evolution of genomic islands by increased establishment probability of linked alleles. *Molecular Ecology*, 25(11), 2542–2558. <https://doi.org/10.1111/mec.13611>

A. Supplementary material: Chapter 2

Appendix 1

Supporting Information Appendix S1 for the submitted manuscript:

The effect of divergent and parallel selection on the genomic landscape of divergence

Ali, H. A., Coulson, T., Clegg, S. M., & Quilodrán, C. S. (2024). The effect of divergent and parallel selection on the genomic landscape of divergence. *Molecular Ecology*, 33(3), e17225.

Questions to: Hisham A.A. Ali. Department of Zoology, University of Oxford.

Email: hisham.a.ali@gmail.com

This vignette provides a description of how to set up *glads* in R and four examples implementing the modifications from this manuscript.

Example 1 : Dominance

Example 2 : Genotype-Environment (GxE) interaction

Example 3 : Frequent dependent selection (FDS)

Example 4 : Variable migration with dominance, GxE interaction and negative FDS

R set up

Within R, the following packages need to be installed:

```
library('ape')
library('ade4')
library('adegenet')
library('pegas')
library('Rcpp')
library('abind')
library('parallel')
library('progress')
library('ggplot2')
library('dplyr')
library('caTools')
```

The *glads* package can be cloned from Github in terminal by running the following commands

```
git clone https://github.com/eriqande/glads
```

```
R CMD INSTALL --no-multiarch --with-keep.source glads
```

After installing *glads*, the library can be loaded into R

```
library('glads')
```

New R and C++ Functions

These are the modified R functions.

The modified `fitness` function

```
fitnessFDS <- function (z, N, b0, b1, b2, b3, d.v, add.loci, FDS,
                        perc, bands, tstatus, yp1, yp2, dimStruct)
{
  ze <- z[, "z"]
  ng <- add.loci
```

```

a <- b0
b <- b1
c <- b2
w <- round(a * exp(-((ze - b * ng)^2)/(2 * (c * ng)^2)) -
           b3 * N + rnorm(N, 0, d.v), 0)

step <- ( ( max(ze) - floor(min(ze)) ) / bands )

t <- ze
zBand <- rep(0,length(t))

for(i in 1:length(t)) {
  zBand[i] <- ceiling((t[i]-floor(min(ze)) )/step) #band each z in order
  zBand[i] <- ifelse( zBand[i] == 0 , 1,zBand[i])
}

fqtab <-as.data.frame(table(zBand))

perc <-perc
FDSfp <- ( fqtab[,2] / max(fqtab[,2]) ) * perc

FDSall<- cbind(fqtab,FDSfp)

zFDS <-rep(0,length(t))

for(i in 1:length(t)) {
  zFDS[i] <- FDSall [zBand[i]== FDSall$zBand , 3]
}

switch(FDS, positive = { # new parameter
  for(i in 1:length(w)) {
    w[i] <- round(w[i] + zFDS[i],0)
  }
},
negative = {
  zFDS<-(perc-zFDS)
  for(i in 1:length(w)) {
    w[i] <- round( w[i] + zFDS[i],0)
  }
}
)
w <- ifelse(w<0,0,w)
w <- ifelse(w>b0,b0,w)

w[is.na(w)] <- 0
return(w)
}

```

The modified `phenotype` function

```
phenotype2 <- function(struct,interact, ...) {  
  G <- struct  
  dims <- dim(G)  
  bvs <- bvs  
  add.loci <- add.loci  
  sex.ratio <- sex.ratio  
  e.v <- e.v  
  inter <- interact  
  rcpp_g2p_map2(G, dims, bvs, add.loci, sex.ratio, e.v,inter)  
}
```

The modified `newborns` function

```
newbornsAll <- function (x,recombination,type, tstatus,yp1, yp2,  
                          dominance, GxE,interact,FDSoption,FDS,perc,bands)  
{  
  struct <- x[[1]]  
  recom.rate <- x[[2]]  
  init.sex <- x[[3]]  
  mutation.rate <- x[[4]]  
  loci.pos <- x[[5]]  
  chromo_mb <- x[[6]]  
  param.z <- x[[7]]  
  param.w <- x[[8]]  
  switch(type, constant = {  
    mean.fitness <- 2  
    n <- nrow(struct)  
    if (n%%2 == 0) {  
      z <- as.data.frame(cbind(sex = sample(rep(0:1, each = n/2)) + 1))  
      rownames(z) = rownames(struct)  
    }  
  })  
}
```

```

else {
neven = floor(n) + floor(n)%%2 - 2
z <- as.data.frame(cbind(sex = c(sample(rep(0:1, each = neven/2)),
                                sample(0:1, 1)) + 1))
rownames(z) = rownames(struct)
}

if (!is.null(init.sex)) {
  z <- as.data.frame(cbind(sex = init.sex))
  rownames(z) = rownames(struct)
}

z$z <- rep(mean.fitness, n)
},
  dynamic = {
sex.ratio <- param.z[[1]]
mean.fitness <- param.w[[1]]
d.d <- param.w[[2]]
n <- nrow(struct)
z <- as.data.frame(cbind(sex = rbinom(n, 1,
sex.ratio)+ 1))
rownames(z) = rownames(struct)
z$z <- round(pmax(rpois(n, lambda = mean.fitness) - d.d * n, 0))
},
  additive = {
sex.ratio <- param.z[[1]]
fitness.pos <- param.z[[2]]
bvs <- param.z[[3]]
add.loci <- param.z[[4]]
e.v <- param.z[[5]]
b0 <- param.w[[1]]
b1 <- param.w[[2]]
b2 <- param.w[[3]]
b3 <- param.w[[4]]
d.v <- param.w[[5]]
n.loci <- param.w[[6]]
struct.fit <- struct[, fitness.pos, ]
struct.fit.dom <- (struct.fit * 0)
bvD. <- dominance

  for (i in 1:dim(struct.fit.dom)[1])
  {
    for (j in 1:dim(struct.fit.dom)[2])
    {
x<- c(bvD[(struct.fit[i,j,1])] , bvD[(struct.fit[i,j,2])])
      if (x[1] > x[2]) {
struct.fit.dom[i,j,1] <- (struct.fit[i,j,1])
struct.fit.dom[i,j,2] <- (struct.fit[i,j,1])
      }
    }
  }
  else if (x[1] == x[2])

```

```

        {
          struct.fit.dom[i,j,1]<- (struct.fit[i,j,1])
          struct.fit.dom[i,j,2]<- (struct.fit[i,j,2])
        } else if (x[1] < x[2])
        {
          struct.fit.dom[i,j,1]<- (struct.fit[i,j,2])
          struct.fit.dom[i,j,2]<- (struct.fit[i,j,2])
        }
      }
    }

    switch(
      GxE,
      on = {
        zs <- phenotype2(struct.fit.dom, bvs, add.loci, sex.ratio,
                        e.v,interact=interact)
      },
      off = {
        zs <- phenotype(struct.fit.dom, bvs, add.loci, sex.ratio,
                        e.v)
      }
    )
    # modifcation of struct.fit

    stop("Invalid GxE type. Current options are 'on' or 'off'")

  )

  dimStruct <-dim(struct)

  switch(
    FDSoption,
    on = {
      fit <- fitnessFDS(zs, dim(zs)[1], b0, b1, b2, b3, d.v,n.loci,FDS,perc,
                      bands,tstatus,yp1,yp2,dimStruct)
    },
    off = {
      fit <- fitness(zs, dim(zs)[1], b0, b1, b2, b3, d.v, n.loci)
    }
  ),

  stop("Invalid FDS type. Current options are 'on' or 'off'")

)

z <- cbind(sex = zs[, "sex"], z = fit)
},

```

```

    custom = {
zs  <- do.call("phenotype", append(list(struct = struct), param.z))
fit <- do.call("fitness", append(list(z = zs[, "z"],
sex = zs[, "sex"], n = dim(zs)[1]), param.w))

z <- cbind(sex = zs[, "sex"], z = fit)
    },

stop("Invalid type of evolution. Current options are 'constant',
'dynamic', 'additive', and 'custom'"))

pairs <- mating(z, struct)
mums  <- pairs[[1]]
dads  <- pairs[[2]]
    if (!is.null(mutation.rate)) {
mums <- mutation(mums, mutation.rate)
dads <- mutation(dads, mutation.rate)

    }

struct.rt <- array(NA, dim(mums))
    switch(recombination, map = {
struct.rt[, , 1] <- rcpp_recombo_seggregate(mums, dim(mums), recom.rate)
struct.rt[, , 2] <- rcpp_recombo_seggregate(dads, dim(dads), recom.rate)
    }, average = {
struct.rt[, , 1] <- rcpp_recombo_seggregate_expo(mums, dim(mums),
loci.pos, chromo_mb, recom.rate)
struct.rt[, , 2] <- rcpp_recombo_seggregate_expo(dads, dim(dads),
loci.pos, chromo_mb, recom.rate)
    },

stop("Invalid recombination type. Current options are 'map' or
'average'"))

return(struct.rt)

}

```

The modified `evolve` function

```

evolveAll<- function (x, time,
                      type = c("constant","dynamic","additive","custom"),

```

```

        recombination = c("map", "average"), recom.rate,
        loci.pos = NULL, chromo_mb = NULL,

        init.sex = NULL,
        migration.rate = NULL,
        mutation.rate = NULL,
        param.z = NULL, param.w = NULL,

        fun = c(phenotype = NULL, fitness = NULL),
        dominance,      # vector representing dominance
        GxE = c("on", "off"), interact,
        FDSoption = c("on", "off"), FDS, perc, bands
    )
    {
        npop <- length(x)
        struct <- x
        if (is.list(x) == F)

stop("Initial populations should be a list of arrays.", call. = F)

        if (recombination == "average" && (is.null(loci.pos)
|| is.null(chromo_mb)))

stop("Loci position or size of the chromosome is missing with no
default for 'average' recombination.", call. = F)

        if ((recombination == "map" && length(recom.rate) != (dim(x[[1]])[2]
-1)) || (recombination == "average" && length(recom.rate) != 1))

stop("Incorrect recombination rate for the type of recombination
('map' or 'average').", call. = F)

        if (type == "custom" && (is.character(fun) == FALSE || length(fun)
!= 2))

stop("One or both custom function names are missing or 'fun' is not
well defined. 'fun' is a character vector of length 2, with the
name of the phenotype and fitness functions e.g. c('phenotype',
'fitness').", call. = F)

        if (!is.null(mutation.rate)) {
            if (sum(sapply(1:npop, function(i) { length(table(x[[i]])) != 2}))
!= 0)
        {
            mutation.rate = NULL

```

```

warning("Mutation rate is ignored for non-biallelic genetic
structures. The genetic structure input should contains integers of
values 1 or 2.", call.=F)
}

else if (sum(sapply(1:npop, function(i) {1 %in% x[[i]] && 2 %in%
x[[i]]
  }) == FALSE) != 0)
{
  mutation.rate = NULL

  warning("Mutation rate is ignored for non-biallelic genetic
structures. The genetic structure input should contains integers of
values 1 or 2.", call.= F)
}
}

if (npop == 1 && !is.null(migration.rate))
{
  migration.rate = NULL

  warning("Migration is ignored for simulations with a single
population.", call. = F)
}

if (!is.null(init.sex) && (length(init.sex) != npop
|| sum(sapply(1:length(x),
function(i) {
  nrow(x[[i]]) != length(init.sex[[i]]))) != 0))

stop("There is not enough females or males assigned to the initial.
populations.", call. = F)

pb <- progress_bar$new(format = " Work in progress [:bar] :percent
eta:      :eta", total = time, clear = FALSE, width = 100)

switch(recombination, map = {
  res      <- list()
  res[["0"]] <- struct
  for (i in 1:time) {
    if (i > 1) {
      init.sex <- NULL

```

```

    }

y <- lapply(1:npop, function(i) {
  append(list(struct[[i]]), list(recom.rate, init.sex[[i]],
    mutation.rate, loci.pos = NULL, chromo_mb = NULL, param.z[[i]],
    param.w[[i]], fun))
})

    )

y1 <- y[[1]]
y2 <- y[[2]]
yp1 <- y1[[1]]
yp2 <- y2[[1]]

out <- lapply(y, newbornsAll,
  recombination = recombination,
  type          = type,
  dominance     = dominance,
  GxE           = GxE, interact=interact,
  FDSoption     = FDSoption, FDS=FDS, perc=perc,
  bands=bands,
  tstatus      = i, yp1=yp1, yp2=yp2
)

if (!is.null(migration.rate)) {
  move <- combn(1:npop, 2)
  for (j in 1:ncol(move)) {
    n1      <- move[1, j]
    n2      <- move[2, j]
    rate1to2 <- migration.rate[i]
    rate2to1 <- migration.rate[i]
    outd     <- migrate(out[[n1]], out[[n2]], rate1to2,
rate2to1)
    out[[n1]] <- outd[[1]]
    out[[n2]] <- outd[[2]]
  }
}
struct <- out

if (i %% 100==0) res[[paste(i)]] <- struct
pb$tick()
}
}, average = {
  res      <- list()
  res[["0"]] <- struct
  for (i in 1:time) {
    if (i > 1) {

```

```

    init.sex <- NULL
  }

y <- lapply(1:npop, function(i) {
  append(list(struct[[i]]), list(recom.rate, init.sex[[i]],
    mutation.rate, loci.pos, chromo_mb, param.z[[i]], param.w[[i]], fun))
})

)

out <- lapply(y,
  newbornsAll,
  recombination = recombination,
  type          = type,
  dominance     = dominance,
  GxE           = GxE, interact=interact,
  FDSoption     = FDSoption, FDS=FDS, perc=perc,
  bands=bands,
  tstatus      = i, yp1=yp1, yp2=yp2 )

  if (!is.null(migration.rate)) {
    move <- combn(1:npop, 2)
    for (j in 1:ncol(move)) {
      n1 <- move[1, j]
      n2 <- move[2, j]
      rate1to2 <- migration.rate[i]
      rate2to1 <- migration.rate[i]
      outd <- migrate(out[[n1]], out[[n2]], rate1to2, rate2to1)
      out[[n1]] <- outd[[1]]
      out[[n2]] <- outd[[2]]
    }
  }
  struct <- out
  if (i %% 100 == 0) res[[paste(i)]] <- struct
  pb$tick()
},
),

stop("Invalid recombination type. Current options are 'map' or
'average'"))

return(res)

}

```

The following modification is in C++

```
cppFunction('
```

```
NumericVector rcpp_g2p_map2(IntegerVector G, IntegerVector dims,  
NumericMatrix bvs, double add_loci, double sex_ratio, double  
e_v, double inter) {
```

```
    int i, l, a, k; // for subscripting individual, locus, gene copy  
    int Ni = dims[0];  
    int Nl = dims[1];  
    int Na = dims[2];  
    double sum;
```

```
    // allocate space to return the result  
    NumericVector ret(Ni, 0.0);  
    NumericMatrix retmat(Ni, 2);
```

```
    for(i=0; i<Ni; i++) {  
        sum = 0.0;  
        for(l=0; l<Nl; l++) {  
            for(a=0; a<Na; a++) {  
                sum += bvs(l, G[a*(Ni*Nl) + l*(Ni) + i] - 1);  
            }  
        }  
        ret[i] = sum - add_loci;  
    }
```

```
    NumericVector envi = rnorm(Ni, 0, e_v);
```

```
    for(k=0; k<Ni; k++) {
```

```
        ret[k] = ret(k) + envi(k) + (inter * envi(k) * ret(k) );
```

```
    }
```

```
    retmat(_, 0) = ret; // make the 1st col of the return matrix be z  
    phenotype  
    retmat(_, 1) = (runif(Ni) < sex_ratio) + 1; // make 2nd col return  
    sex  
    colnames(retmat) = CharacterVector::create("z", "sex");
```

```
    return(retmat);  
}
```

```
')
```

Parameters

These are the parameters needed for the examples below

We use the base random seed function as a pseudo-random generator for reproducibility

```
set.seed(1)
```

We then build our *glads* populations using the `initial.struct` function included in the *glads* package

```
initial.population.size=400 #Initial population size
n.loci=300                  #Number of loci
n.alleles.per.locus=20      #Number of alleles

start.1 <- start.2 <-
initial.struct(initial.population.size,n.loci,n.alleles.per.locus)
start <- list(start.1, start.1) #Initial populations
```

The selected loci are identified by the following code

```
loci.pos=1:n.loci          # Loci position ;creates a sequence 1 to 300
add.loci=50                 # Number of additive loci
fitness.pos <- 125:(125+(add.loci-1)) # Additive loci position
```

The breeding values are then given

```
bvs = t(                                     #Breeding value of additive loci
  array(                                     #sets a matrix
    seq(0,1, length = n.alleles.per.locus), #20 values from 0 to 1
    c(n.alleles.per.locus, add.loci)         #20 by 50; matrix dimensions
  )                                           #columns are repeat of value
)
```

The fitness function and demographic parameters are assigned to each population

```
sex.ratio <- 0.5
e.v=0.01          #Stochastic environmental variant
d.v=1             #Stochastic demographic variant
n.gens=100        #Number of generations
recom.map= c(rep(0.5, 299)) #Recombination map; 0.5 repeated 299 times
Linked<-c(60:69, 150:159, 230:239) # Linked loci
recom.map[Linked]<-0.0001 #Recombination map with linked loci
param.w1 <- list(b0 = 6,b1 = 0.25, b2 = 0.5, b3 = 0.01, d.v = d.v,
```

```

        add.loci = add.loci) # fitness function parameters
param.w2 <- list(b0 = 6, b1 = 0.25, b2 = 0.5, b3 = 0.01, d.v = d.v,
               add.loci = add.loci) # same for parallel selection

param.z1 <- list(
  sex.ratio,      #0.5
  fitness.pos,    #125:174 length=50
  bvs,            #matrix 20vals, rep.x50
  add.loci,       #50
  e.v             #Stochastic environmental variant,0.01 #  $\sigma$ -env eq.1
)

param.z2 <- param.z1

```

Example 1: Dominance

This example simulates genetic dominance using *glads*, which took 2.3 seconds in Rstudio in a MAC OS 15.1 with 64 bits 4.4 GHz

```

example.dominance <- evolveAll(
  x = start,                # initial genetic structure of populations
  time = n.gens,            # number of generations;
  type = "additive",        # type of simulated evolution
  recombination = "map",    # type of recombination
  recom.rate = recom.map,
  param.z = list(param.z1, param.z2), # phenotype function parameters
  param.w = list(param.w1, param.w2), # fitness function parameters
  dominance = c(1:20),      # dominance hierarchy for alleles
  GxE='off',
  FDSoption='off',
  migration.rate= rep(0,n.gens) # no migration
)

```

The output of the *glads* model now needs to be analysed to generate the genomic landscape of divergence

First, we need to convert *glads* to the *pegas* format

```

example.loci.dominance <- struct2pegas(example.dominance[[2]])

```

Then we need to calculate F_{ST} between the populations

```
FST.dominance <- pegas::Fst(example.loci.dominance)
```

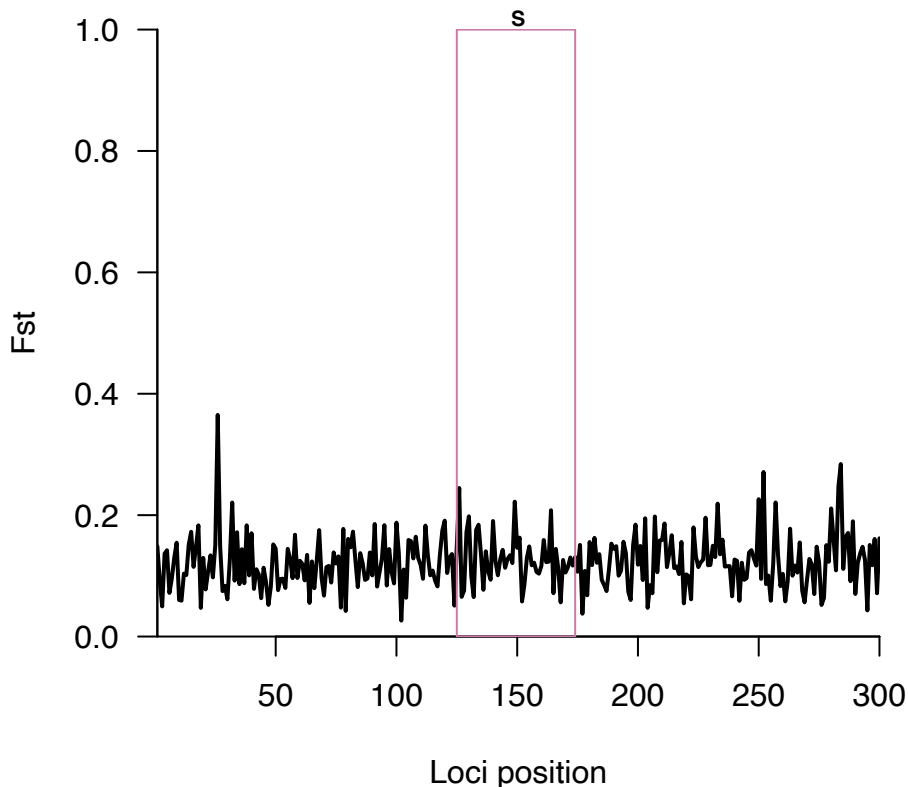
```
avFst.dominance <- sum(FST.dominance)/nrow(FST.dominance) # average  
Fst across all loci
```

An example script for producing a figure of the average F_{ST} is provided below

```
plot(loci.pos,  
     FST.dominance[, "Fst"],  
     pch=16, bty="l",  
     xlab="Loci position", ylab="Fst",  
     ylim=c(0,1),  
     type="l",  
     xaxs="i", yaxs="i",  
     las=1, lwd=2  
     )
```

A box can be added to show the loci under selection

```
rect( xleft=125,xright = 174,ybottom=0,ytop=1,title(main='S',line  
=0.1,cex.main=0.7),border='#CC79A7') # show selected loci
```



Example 2 : Genotype-Environment (GxE) interaction

This example simulates genotype-environment (GxE) interaction using *glads*, which took 2.2 seconds in Rstudio in a MAC OS 15.1 with 64 bits 4.4 GHz

```
example.GxE <- evolveAll(
  x = start,                # initial genetic structure of populations
  time = n.gens,            # number of generations
  type = "additive",        # type of simulated evolution
  recombination = "map",    # type of recombination
  recom.rate = recom.map,
  param.z = list(param.z1, param.z2), # phenotype function parameters
  param.w = list(param.w1, param.w2), # fitness function parameters
  dominance = rep(1,n.alleles.per.locus), # all alleles are equal in
  hierarchy, no allele has dominance over another
  GxE='on',                # genotype - environment interaction
  interact=10,              # interaction strength, G in eq.6 of manuscript
  FDSoption='off',
  migration.rate= rep(0,n.gens)
)
```

The output of the *glads* model now needs to be analysed to generate the genomic landscape of divergence

First, we need to convert *glads* to the *pegas* format

```
example.loci.GxE <- struct2pegas(example.GxE[[2]]) # convert to pegas
format
```

Then we need to calculate F_{ST} between the populations

```
FST.GxE <- pegas::Fst(example.loci.GxE)
```

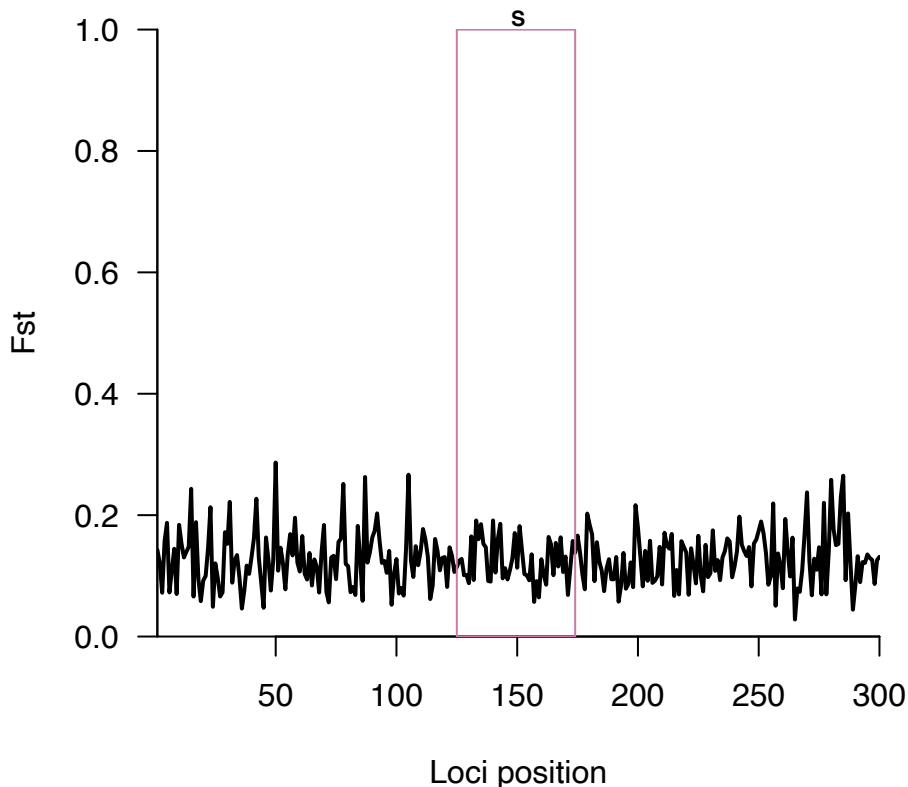
```
avFst.GE<- sum(FST.GxE)/nrow(FST.GxE) # average Fst across all loci
```

An example script for producing a figure of the average F_{ST} is provided below

```
plot(loci.pos,
     FST.GxE[, "Fst"],
     pch=16, bty="l",
     xlab="Loci position", ylab="Fst",
     ylim=c(0,1),
     type="l",
     xaxs="i", yaxs="i",
     las=1, lwd=2
)
```

A box can be added to show the loci under selection

```
rect( xleft=125,xright = 174,ybottom=0,ytop=1,title(main='S',line
=0.1,cex.main=0.7),border='#CC79A7') # show selected loci
```



Example 3 : Frequency dependent selection (FDS)

This example simulates frequency dependent (FDS) selection using *glads*. FDS can be negative or positive, and in this example, it is simulated as negative (NFDS). The simulation took 3.3 seconds in Rstudio in a MAC OS 15.1 with 64 bits 4.4 GHz

```
example.FDS <- evolveAll(
  x = start,                                # initial genetic structure of populations
  time = n.gens,                            # number of generations
  type = "additive",                        # type of simulated evolution
  recombination = "map",                    # type of recombination
  recom.rate = recom.map,
  param.z = list(param.z1, param.z2),      # phenotype function parameters
  param.w = list(param.w1, param.w2),      # fitness function parameters
  dominance = rep(1, n.alleles.per.locus),
  GxE='off',
  FDSoption='on',                          # introduce frequency-dependent selection
  FDS='negative', # 'negative' or 'positive' frequency-dependent sel.
  perc=2,                                  # F in eq.3-5 of manuscript
  bands=10,                                # K in eq.3-5 of manuscript
  migration.rate= rep(0, n.gens)
```

```
)
```

The output of the *glads* model now needs to be analysed to generate the genomic landscape of divergence

First, we need to convert *glads* to the *pegas* format

```
example.loci.FDS <- struct2pegas(example.FDS[[2]])
```

```
FST.FDS <- pegas::Fst(example.loci.FDS) # calculate Fst between  
populations
```

Then we need to calculate F_{ST} between the populations

```
FST.FDS <- pegas::Fst(example.loci.FDS)
```

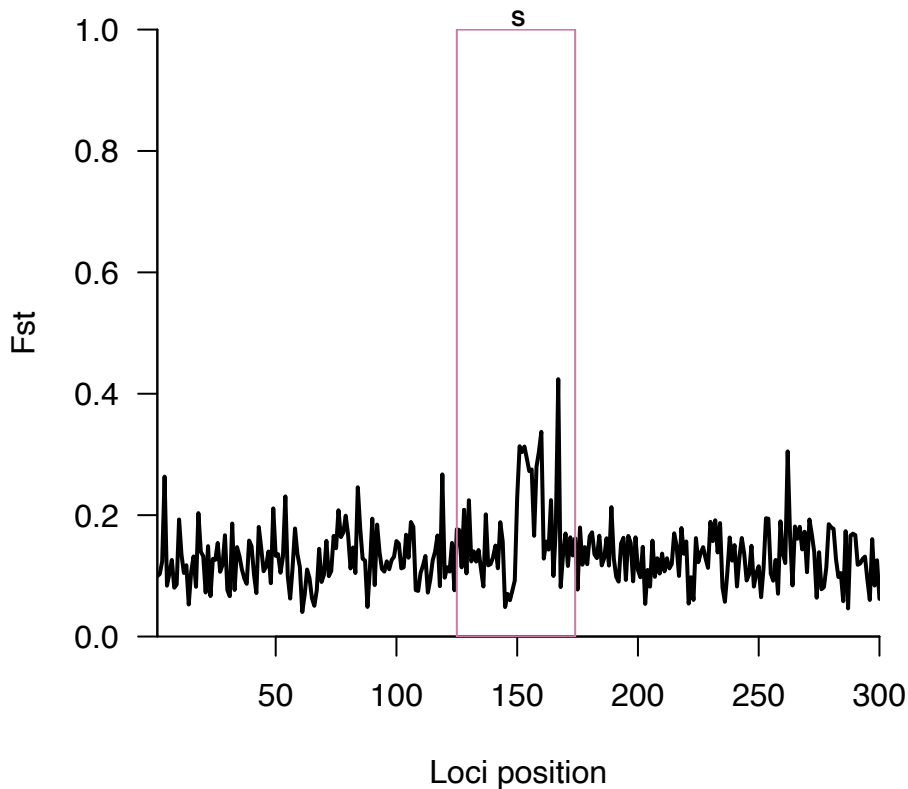
```
avFst.FDS <- sum(FST.FDS)/nrow(FST.FDS) # average Fst across all loci
```

An example script for producing a figure of the average F_{ST} is provided below

```
plot(loci.pos,  
     FST.FDS[, "Fst"],  
     pch=16, bty="l",  
     xlab="Loci position", ylab="Fst",  
     ylim=c(0,1),  
     type="l",  
     xaxs="i", yaxs="i",  
     las=1, lwd=2  
)
```

A box can be added to show the loci under selection

```
rect( xleft=125,xright = 174,ybottom=0,ytop=1,title(main='S',line  
=0.1,cex.main=0.7),border='#CC79A7') # show selected loci
```



Example 4 : Variable migration with dominance, GxE interaction, and negative FDS

Here, in *glads*, the following modifications are simulated concurrently:

- (1) dominance (2) GxE interaction (3) frequency-dependent selection (4) variable migration rate

The simulation took 3.4 seconds in Rstudio in a MAC OS 15.1 with 64 bits 4.4 GHz

```
example.all <- evolveAll(
  x = start,                                # initial genetic structure of populations
  time = n.gens,                            # number of generations
  type = "additive",                        # type of simulated evolution
  recombination = "map",                    # type of recombination
  recom.rate = recom.map,
  param.z = list(param.z1, param.z2),      # phenotype function parameters
  param.w = list(param.w1, param.w2),      # fitness function parameters
  dominance = c(1,2,3,4,5,6,5,4,3,2,2,3,4,5,6,5,4,3,2,1), #dom hierarchy
  GxE='on',                                # genotype - environment interaction
  interact=10,                              # interaction strength, G in eq.6 of manuscript
  FDSoption='on',
```

```

FDS='negative',          # negative frequency-dependent selection
perc=2,                  # F in eq.3-5 of manuscript
bands=10,                # K in eq.3-5 of manuscript
migration.rate=c(        # migration rate for each of the 100 generations
  rep(0,10),

  seq(0.01,0.1,0.01),rep(0.1,10),rev(seq(0.01,0.1,0.01)),
  rep(0,20),

  seq(0.01,0.1,0.01),rep(0.1,10),rev(seq(0.01,0.1,0.01)),
  rep(0,10))

)

```

The output of the *glads* model now needs to be analysed to generate the genomic landscape of divergence

First, we need to convert *glads* to the *pegas* format

```
example.loci.all <- struct2pegas(example.all[[2]])
```

Then we need to calculate F_{ST} between the populations

```
FST.all <- pegas::Fst(example.loci.all)
```

```
avFst.all <- sum(FST.all)/nrow(FST.all) # average Fst across all loci
```

An example script for producing a figure of the average F_{ST} is provided below

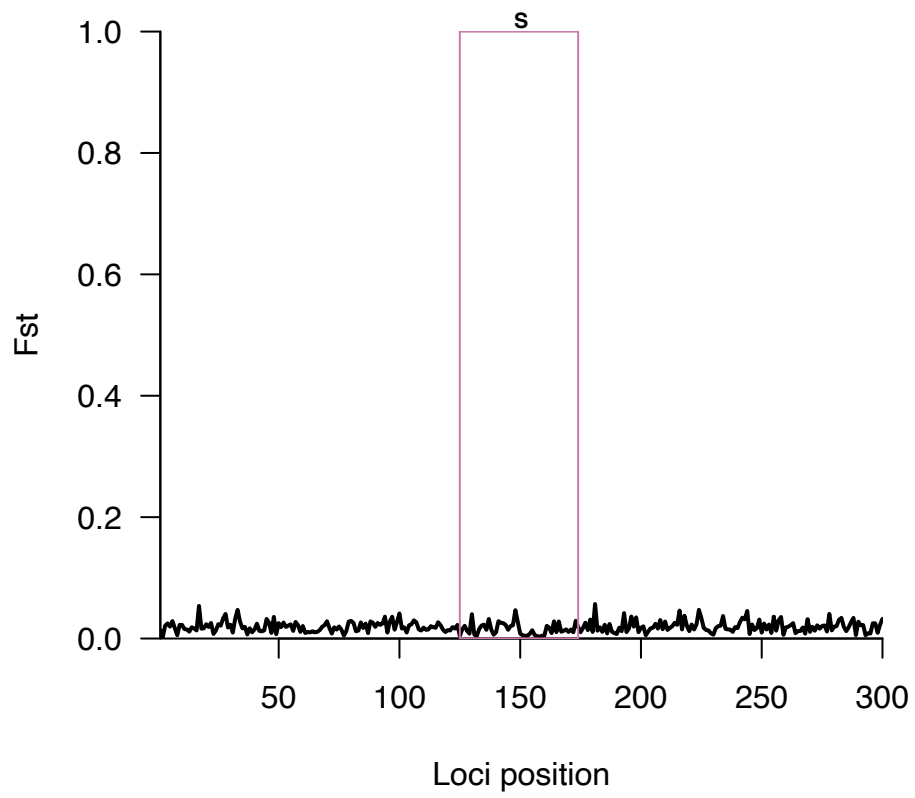
```

plot(loci.pos,
     FST.all[, "Fst"],
     pch=16, bty="l",
     xlab="Loci position", ylab="Fst",
     ylim=c(0,1),
     type="l",
     xaxs="i", yaxs="i",
     las=1, lwd=2
)

```

A box can be added to show the loci under selection

```
rect( xleft=125,xright = 174,ybottom=0,ytop=1,title(main='S',line
=0.1,cex.main=0.7),border='#CC79A7') # show selected loci
```



B. Supplementary material: Chapter 3

Appendix 1

Supporting Information Appendix S1 for the manuscript:

The effect of Dobzhansky-Muller incompatibilities on the genomic landscape of divergence under parallel and divergent selection

Questions to: Hisham A.A. Ali. Department of Zoology, University of Oxford.

Email: hisham.a.ali@gmail.com

This vignette provides a description of how to set up *glads* in R and four examples implementing the modifications from this manuscript.

Example 1 : Epistasis probability method

Example 2 : Epistasis map method

Example 3 : gen02glads

Example 4 : Positive epistasis

R set up

Within R, the following packages need to be installed:

```
library('ape')
library('ade4')
library('adegenet')
library('pegas')
library('Rcpp')
library('abind')
library('parallel')
library('progress')
library('ggplot2')
library('dplyr')
library('caTools')
```

The *glads* package can be cloned from Github in terminal by running the following commands

```
git clone https://github.com/eriqande/glads
```

```
R CMD INSTALL --no-multiarch --with-keep.source glads
```

After installing *glads*, the library can be loaded into R

```
library('glads')
```

New R and C++ Functions

These are the modified R functions.

The modified `evolve` function

```
evolveM <-
  function (x, time,
            type = c("constant", "dynamic", "additive", "custom"),
            recombination = c("map", "average"),
```

```

    recom.rate,
    loci.pos = NULL, chromo_mb = NULL,

    init.sex = NULL,

    migration.rate = NULL,
    mutation.rate = NULL,

    param.z = NULL, param.w = NULL,

    fun = c(phenotype = NULL, fitness = NULL),
    DMI.type= c("probability","epistasis"),epistasis.map,pDMI,

    h,
    trait1.loci,trait2.loci
  )
}

npop <- length(x)
struct <- x
if (is.list(x) == F)

stop("Initial populations should be a list of arrays.", call. = F)

if (recombination == "average" && (is.null(loci.pos)
  || is.null(chromo_mb)))

stop("Loci position or size of the chromosome is missing with no
  default for 'average' recombination.",call. = F)

if ((recombination == "map" && length(recom.rate) != (dim(x[[1]])[2]
- 1)) || (recombination == "average" && length(recom.rate) != 1))

stop("Incorrect recombination rate for the type of recombination ('map
  ' or 'average').",call. = F)

if (type == "custom" && (is.character(fun) == FALSE || length(fun)
  != 2))

stop("One or both custom function names are missing or 'fun' is not we
ll defined. 'fun' is a character vector of length 2, with the name of
the phenotype and fitness functions e.g. c('phenotype', 'fitness').",
call. = F)

```

```

if (!is.null(mutation.rate)) {
  if (sum(sapply(1:npop, function(i) {
    length(table(x[[i]])) != 2
  })) != 0) {
    mutation.rate = NULL

warning("Mutation rate is ignored for non-biallelic genetic structures
. The genetic structure input should contains integers of values 1 or
2.", call. = F)
  }

  else if (sum(sapply(1:npop, function(i) {
    1 %in% x[[i]] && 2 %in% x[[i]]
  })) == FALSE) != 0) {
    mutation.rate = NULL

warning("Mutation rate is ignored for non-biallelic genetic structures
. The genetic structure input should contains integers of values 1 or
2.", call. = F)

  }
}
if (npop == 1 && !is.null(migration.rate)) {
  migration.rate = NULL

warning("Migration is ignored for simulations with a single population
.", call. = F)
}

if (!is.null(migration.rate)) {
  if (length(migration.rate) == 1) {
    disp = matrix(rep(migration.rate, npop * npop), nrow = npop,
                  ncol = npop, byrow = TRUE)
    diag(disp) <- 1
  }

  else {
    disp = migration.rate
    diag(disp) <- 1
    if (sum(dim(migration.rate) != npop) != 0)

stop("Migration should be a single value or an square matrix of order
equal to the number of populations.", call. = F)
  }
}

```

```

    }
  }

  if (!is.null(init.sex) && (length(init.sex) != npop || sum(sapply(
    1:length(x),

function(i) {
  nrow(x[[i]]) != length(init.sex[[i]]))) != 0))

  stop("There is not enough females or males assigned to the initial pop
      ulations.", call. = F)

  pb <- progress_bar$new(format = " Work in progress [:bar] :percent eta
      : :eta", total = time, clear = FALSE, width = 100)

  switch(recombination, map = {
    res <- list()
    for (i in 1:time) {
      if (i > 1) {
        init.sex <- NULL
      }

      y <- lapply(1:npop, function(i) {
append(list(struct[[i]]), list( recom.rate, init.sex[[i]],
      mutation.rate,

      loci.pos = NULL,

      chromo_mb = NULL,
      param.z[[i]],

      param.w[[i]],

      fun,

      paste('pop',i,sep = ""))) # change

      })

    out <- lapply( y, newbornsM,

      recombination = recombination,
      type = type,DMI.type=DMI.type,

      epistasis.map=epistasis.map,pDMI,

      h, trait1.loci,trait2.loci

```

```

    )

    if (!is.null(migration.rate)) {
      move <- combn(1:npop, 2)
      for (j in 1:ncol(move)) {
        n1 <- move[1, j]
        n2 <- move[2, j]
        rate1to2 <- disp[n1, n2]
        rate2to1 <- disp[n2, n1]
        outd <- migrate(out[[n1]], out[[n2]], rate1to2,
                        rate2to1)
        out[[n1]] <- outd[[1]]
        out[[n2]] <- outd[[2]]
      }
    }

    struct <- out
    if (i%%time == 0) res <- struct
    pb$tick()
  },

  average = {
    res <- list()
    for (i in 1:time) {
      if (i > 1) {
        init.sex <- NULL
      }

      y <- lapply(1:npop, function(i) {
        append(list(struct[[i]]),

                list(recom.rate,

                     init.sex[[i]],
                     mutation.rate,

                     loci.pos,

                     chromo_mb,

                     param.z[[i]], param.w[[i]],

                     fun,paste('pop',i,sep = "")))

      })
    }
  }
}

```

```

out <- lapply(y, newbornsM,

              recombination = recombination,
              type = type,

              DMI.type=DMI.type, epistasis.map=epistasis.map,pDMI,

              h,trait1.loci,trait2.loci)

if (!is.null(migration.rate)) {
  move <- combn(1:npop, 2)
  for (j in 1:ncol(move)) {
    n1 <- move[1, j]
    n2 <- move[2, j]
    rate1to2 <- disp[n1, n2]
    rate2to1 <- disp[n2, n1]
    outd <- migrate(out[[n1]], out[[n2]], rate1to2,
                    rate2to1)
    out[[n1]] <- outd[[1]]
    out[[n2]] <- outd[[2]]
  }
}

struct <- out
if (i%%time == 0) res <- struct
pb$tick()
},

stop("Invalid recombination type. Current options are 'map' or
                                           'average'"))

return(res)

}

```

The modified `newborns` function

```

newbornsM <-
  function (x, recombination, type,DMI.type,epistasis.map,pDMI,h,

            trait1.loci,trait2.loci)
  {
    struct      <- x[[1]]
    recom.rate  <- x[[2]]
    init.sex    <- x[[3]]
    mutation.rate <- x[[4]]
    loci.pos    <- x[[5]]

```

```

chromo_mb      <- x[[6]]
param.z        <- x[[7]]
param.w        <- x[[8]]
fun            <- x[[9]]
popID          <- x[[10]]

switch(type,
  constant = {
    mean.fitness <- 2

    if (!is.null(mutation.rate)) {

      struct<- mutationM(struct,mutation.rate,popID,pDMI)

    }

    n <- nrow(struct)
    if (n%%2 == 0) {
      z <- as.data.frame(cbind(sex = sample(rep(0:1, each = n / 2)) + 1))
      rownames(z) = rownames(struct)
    } else {

      neven = floor(n) + floor(n)%%2 - 2

      z <- as.data.frame(cbind(sex = c(
        sample(rep(0:1, each = neven/2) ), sample(0:1, 1)) + 1))

      rownames(z) = rownames(struct)
    }

    if (!is.null(init.sex)) {
      z <- as.data.frame(cbind(sex = init.sex))
      rownames(z) = rownames(struct)
    }

    z$z <- rep(mean.fitness, n)
  },

  dynamic = {
    sex.ratio <- param.z[[1]]
    mean.fitness <- param.w[[1]]
    d.d <- param.w[[2]]
  }

```

```

    if (!is.null(mutation.rate)) {

        struct<- mutationM(struct,mutation.rate,popID,pDMI)

    }

n <- nrow(struct)
z <- as.data.frame(cbind(sex = rbinom(n, 1, sex.ratio) + 1))
    rownames(z) = rownames(struct)
z$z <- round(pmax(rpois(n, lambda = mean.fitness) - d.d * n, 0))
},
    additive = {
        sex.ratio    <- param.z[[1]]
        fitness.pos  <- param.z[[2]]
        bvs          <- param.z[[3]]
        add.loci     <- param.z[[4]]
        e.v          <- param.z[[5]]
        b0           <- param.w[[1]]
        b1.trait1    <- param.w[[2]]
        b1.trait2    <- param.w[[3]]
        b2.trait1    <- param.w[[4]]
        b2.trait2    <- param.w[[5]]
        b3           <- param.w[[6]]
        d.v          <- param.w[[7]]
        n.loci       <- param.w[[8]]
        trait.correlation <- param.w[[9]]

    if (!is.null(mutation.rate)) {

        struct<- mutationM(struct,mutation.rate,popID,pDMI)

    }

    struct.fit.trait1 <- struct[, trait1.loci, ]
    struct.fit.trait2 <- struct[, trait2.loci, ]
    trait1.bvs<- match(trait1.loci,fitness.pos)
    trait2.bvs<- match(trait2.loci,fitness.pos)

zs.trait1 <- phenotypeM(struct.fit.trait1, bvs[trait1.bvs,], add.loci,
sex.ratio, e.v,struct,DMI.type,epistasis.map,h,b1.trait1,trait1.loci)

zs.trait2 <- phenotypeM(struct.fit.trait2, bvs[trait2.bvs,], add.loci,
sex.ratio, e.v,struct,DMI.type,epistasis.map,h,b1.trait2,trait2.loci)

ploid<-dim(struct)[3]

```

```

        fit <- fitnessM( zs.trait1,zs.trait2,      # trait fitness
                        dim(zs.trait1)[1], # number of indiv.=
                        b0,
                        b1.trait1,b1.trait2, # b1 for each trait
                        b2.trait1,b2.trait2, # b2 for each trait
                        b3, d.v,
                        trait.correlation,
                        ploid,trait1.loci,trait2.loci
                    )

        z <- cbind(sex = zs.trait1[, "sex"], z = fit)

# sex generated for each pheno , we will just use from trait1
# as both are random assignments based on sex.ratio

    },
    custom = {

        if (!is.null(mutation.rate)) {

            struct<- mutationM(struct,mutation.rate,popID,pDMI)

        }

zs <- do.call(fun[1], append(list(struct = struct), param.z))
fit <- do.call(fun[2], append(list(z = zs[, "z"], sex = zs[, "sex"],
                                n = dim(zs)[1]), param.w))

z <- cbind(sex = zs[, "sex"], z = fit)
    },

stop("Invalid type of evolution. Current options are 'constant', 'dynamic', 'additive', and 'custom'"))

    pairs <- mating(z, struct)
    mums <- pairs[[1]]
    dads <- pairs[[2]]

    struct.rt <- array(NA, dim(mums))
    switch(recombination, map = {

struct.rt[, , 1] <-rcpp_recombo_segregate(mums, dim(mums), recom.rate)
struct.rt[, , 2] <-rcpp_recombo_segregate(dads, dim(dads), recom.rate)

```

```

    }, average = {

struct.rt[, , 1] <- rcpp_recombo_seggregate_expo(mums, dim(mums),
                                                loci.pos, chromo_mb, recom.rate)

struct.rt[, , 2] <- rcpp_recombo_seggregate_expo(dads, dim(dads),
                                                loci.pos, chromo_mb, recom.rate)
    },
stop("Invalid recombination type. Current options are 'map' or 'average'"))

    return(struct.rt)
}

```

The modified `mutation` function

```

mutationM <-
function (G, mutation.rate, popID, pDMI)
{
  n <- dim(G)[1]
  nl <- dim(G)[2]
  nmut <- rbinom(1, (nl * n * 2), mutation.rate) #mutation total num
  DMIInmut<- rbinom(1,nmut,pDMI) # how many mutations are DMI type
  Nnmut<- nmut - DMIInmut # non-DMI mutation number
  Npos <- sample(1:(nl * n * 2), Nnmut) # non-DMI mutation position
  DMIPos <- sample(1:(nl * n * 2), DMIInmut) # DMI mutation position

  switch(
    popID,
    pop1 = {
      G[Npos] <- G[Npos]%%2 + 3 # 1&2 -> 4&3
      G[DMIPos] <- G[DMIPos]%%2 + 5 # 1&2 -> 6&5
    },

    pop2 = {
      G[Npos] <- G[Npos]%%2 + 7 # 1&2 -> 8&7
      G[DMIPos] <- G[DMIPos]%%2 + 9 # 1&2 -> 10&9
    }
  )

  return(G)
}

```

The modified `phenotype` function

```
phenotypeM<-
  function (G, bvs, add.loci, sex.ratio, e.v,G.all,DMI.type,epistasis.
map=epistasis.map,h,b1,trait.position)
  {
    dims <- dim(G)
    b1<-as.numeric(b1)

# not all loci used for DMI calculation, only for loci for that trait

    switch(DMI.type,
      probability={

        nDMIpA<-rep(0,nrow(G))
        nDMIpB<-rep(0,nrow(G))
        incompat<-rep(0,nrow(G))
        b1<-as.numeric(b1)

        for (i in 1:nrow(G)) {

# DMI incompatibilities only arise with migration

          nDMIpA[i]<- length( G[i,,] [ G[i,,] == 5 | G[i,,] == 6 ] )
# how many 5|6 DMI
          nDMIpB[i]<- length( G[i,,] [ G[i,,] == 9 | G[i,,] == 10 ] )
# how many 10|9 DMI

          incompat[i]<- min(nDMIpA[i] , nDMIpB[i] )

        }

        fitDEC<-rep(0,length(incompat))

        for (i in 1:length(incompat)) { # change
          fitDEC[i]<- h(incompat[i])
        }

      },
      epistasis = {

        incompat<- rep(0,nrow(G.all)) # length is number of ind.
        fitDEC<-rep(0,nrow(G.all))

        for (i in 1:nrow(G.all)) { # for each ind.

          DMIscoreMAT<-matrix(data=rep(0),nrow=ncol(G.all),ncol=ncol(G.all))
# non-used loci remain as 0
```

```

        for (j in trait.position) {      # number of loci

            for (k in trait.position) {   # number of loci

                if(epistasis.map[j,k]== 1){

                    # when incompatibility exists between j and k

                    if( G.all[i,j,1] == 5 | G.all[i,j,1] == 6 | G.all[i,j,2] == 5 |
G.all[i,j,2] == 6 ) { # loci j is 5|6

                        if( G.all[i,k,1] == 9 | G.all[i,k,1] == 10 | G.all[i,k,2] == 9 |
G.all[i,k,2] == 10 ) { # loci k is 9|10

                            DMIScoreMAT[j,k] <- 1      # non-used loci remain as 0

                            if ( length(
G.all[i,c(k,j),] [ G.all[i,c(k,j),] == 5 | G.all[i,c(k,j),] == 6 ]
)
                                + length(
G.all[i,c(k,j),] [ G.all[i,c(k,j),] == 9 | G.all[i,c(k,j),] == 10 ]
) == 4 )
                                {

                                    DMIScoreMAT[j,k] <- 2 # non-used loci remain as 0
                                }

                            }

                        }else if(

G.all[i,j,1] == 9 | G.all[i,j,1] == 10 |
G.all[i,j,2] == 9 | G.all[i,j,2] == 10 ){ # loci j is 9|10

                            if(

G.all[i,k,1] == 5 | G.all[i,k,1] == 6 |
G.all[i,k,2] == 5 | G.all[i,k,2] == 6 ) { # loci k is 5|6

                                DMIScoreMAT[j,k] <- 1      # non-used loci remain as 0

                            if ( length(

```

```

G.all[i,c(k,j),] [ G.all[i,c(k,j),] == 5 | G.all[i,c(k,j),] == 6 ]
    + length(
G.all[i,c(k,j),] [ G.all[i,c(k,j),] == 9 | G.all[i,c(k,j),] == 10
]    ) == 4 ){
    DMIscoreMAT[j,k] <- 2 # non-used loci remain as 0
    }

    }

    }
}
}

incompat[i]<-sum(DMIscoreMAT)

fitDEC[i]<- h(incompat[i])
}

},

stop("Invalid type of DMI. Current options are 'probability' and 'epi
stasis'")

)

rcpp_g2p_mapM(G, dims, bvs, add.loci, sex.ratio, e.v,b1,fitDEC)
}

```

The modified `fitness` function

```

fitnessM <-
  function( zs.trait1,zs.trait2,
            N,
            b0,
            b1.trait1,b1.trait2,
            b2.trait1,b2.trait2,
            b3, d.v,
            trait.correlation,
            ploid,trait1.loci,trait2.loci
            )

```

```

{

  ze.trait1 <- zs.trait1[, "z"]
  ze.trait2 <- zs.trait2[, "z"]

  a <- b0

  b1t1 <- b1.trait1
  b1t2 <- b1.trait2
  b2t1 <- b2.trait1
  b2t2 <- b2.trait2

  tc<- trait.correlation

t1dev<- ( ( ze.trait1) - (b1t1*ploid*length(trait1.loci)) ) / (b2t1*
                                ploid*length(trait1.loci))  )
t2dev<- ( ( ze.trait2) - (b1t2*ploid*length(trait2.loci)) ) / (b2t2*
                                ploid*length(trait2.loci))  )

corr<- ( - 0.5 * (1- tc^2) )

w <- round( ( a * exp( corr * ( ( t1dev^2) - ( 2 * tc * t1dev *
                                t2dev ) + ( t2dev^2 ) ) ) ) - (b3*N) + rnorm(N, 0, d.v) ,0)

w <- ifelse(w < 0, 0, w)
  return(w)

}

```

The following modification is in C++

```

cppFunction('

NumericVector rcpp_g2p_mapM(IntegerVector G, IntegerVector dims, NumericMatrix bvs, double add_loci, double sex_ratio, double e_v, double b1, NumericVector fitDEC) {

  int i,l,a,k; // for subscripting individual, locus, gene copy then
k
  int Ni = dims[0];
  int Nl = dims[1];
  int Na = dims[2];
  double sum;

```

```

// allocate space to return the result
NumericVector ret(Ni, 0.0);
NumericMatrix retmat(Ni, 2);

for(i=0; i<Ni; i++) {
  sum = 0.0;
  for(l=0; l<Nl; l++) {
    for(a=0; a<Na; a++) {
      sum += bvs(l, G[a*(Ni*Nl) + l*(Ni) + i] - 1);
    }
  }
  ret[i] = sum - add_loci;
}

NumericVector envi = rnorm(Ni, 0, e_v);

for(k=0; k<Ni; k++) {

ret[k] = ret[k] + envi[k] + ( ( ( ( ret[k] + envi[k] ) - b1 ) / abs
(( ( ret[k] + envi[k] ) - b1 ) ) ) *( -1 * fitDEC[k] ) );

}

retmat(_, 0) = ret;

// make the first column of the return matrix be z, the phenotype
retmat(_, 1) = (runif(Ni) < sex_ratio) + 1;

// make the second column of the return be sex

colnames(retmat) = CharacterVector::create("z", "sex");

return(retmat);
}

')

```

Geno2glads

```

geno2glads<- function(seqData, # seqdata = .geno file of sequence data
                        SEED,   #random number generator seed
                        ini.ind) { # pop. size for glads

  ##~~~~~##
  ### format .geno file ###

```

```

##~~~~~##

Geno.t<-t(seqData[,-1])      # trim to get rid of row numbering
Geno.t[1,1] <- "ID"
colnames(Geno.t) <- Geno.t[1,]
Geno.t<-Geno.t[-1,]
rownames(Geno.t) <- Geno.t[,1]
Geno.t<-Geno.t[, -1]        # reshaping data

n1<-apply(Geno.t,1:2,function(i) strsplit(i,split='')[[1]][1])
# split diploid sets into two
n2<-apply(Geno.t,1:2,function(i) strsplit(i,split='')[[1]][2])

##~~~~~##
## make bases numeric ###
##~~~~~##

n1[n1 %in% "A"] <- 1
n1[n1 %in% "C"] <- 2
n1[n1 %in% "G"] <- 3
n1[n1 %in% "T"] <- 4
n1[n1 %in% "N"] <- NA

n1<-apply(n1,2, as.numeric)
rownames(n1)<-rownames(n2)

n2[n2 %in% "A"] <- 1
n2[n2 %in% "C"] <- 2
n2[n2 %in% "G"] <- 3
n2[n2 %in% "T"] <- 4
n2[n2 %in% "N"] <- NA

n2<-apply(n2,2, as.numeric)
rownames(n2)<-rownames(n1)

##~~~~~##
## data structure ###
##~~~~~##

struct.t0<-list(n1,n2)      # put n1 and n2 together in list, length=2

L<-list(struct.t0[[1]], struct.t0[[2]])

struct.t00<-array(unlist(L),
                  dim = c(nrow(L[[1]]), # dim is that of n1 and n2
                          ncol(L[[1]]),

```

```

                                length(L)),      # 2
                                dimnames=list(rownames(L[[1]]),colnames(L[[1]]), 1:2 ))

dat0<-as.data.frame(rbind(struct.t0[[1]],struct.t0[[2]]) )

# n1 and n2 in data frame

rownames(dat0)<-1:nrow(dat0)

##~~~~~##
## generating population based on data ###
## remove na.s ##
## create counts table ##
## create artificial pop. ##
##~~~~~##

snp<-lapply(1:ncol(dat0), function(i){ table(dat0[[i]], exclude=NA) })
# remove na from data;creates a count table

set.seed(SEED)

Pop<-sapply(1:ini.ind, function(i){

sapply(1:ncol(struct.t0[[1]]), function(j)

    { # ncol(n1)=2533;snp number

n1<-sample(names(snp[[j]]),1,T, prob=snp[[j]])
#using weightings to sample; 1 sample; this will be one of the alleles

n2<-sample(names(snp[[j]]),1,T, prob=snp[[j]])
# weighted by occurrence based on count(done by table() to give SNP)

return(list(n1, n2))
# we are doing this twice, 2533 loci for 400 ind.
    })      # ; so we get 5066 loci bases for 400 ind.

}) # so we get 400 ind. with 2 copies at 2533 loci; genome size=5066

##~~~~~##
## modify to evolve format ###
##~~~~~##

Pop<-t(Pop)      # data flipped; dimensions are now reversed
even<-seq_len(ncol(Pop)) %% 2 ==0
# for 400 ind. string of true (even) and false(odd)

n1<-Pop[, even]      # calls even col.s of n1, ind.

```

```

n2<-Pop[, !even] # calls odd col.s of n1, ind.

n1<-apply(n1,2, as.numeric) # set to columns, coerce numeric
n2<-apply(n2,2, as.numeric)
L<-list(n1,n2) # string together, 2 sets of 5066,200

startOUT <- array(unlist(L), # all L in string
                  dim = c(nrow(L[[1]]), # setting dimensions for array
                          ncol(L[[1]]),
                          length(L)), # 2 ; dim(start)=400,2533,2
                  dimnames=list(1:ini.ind,colnames(dat0), 1:2 )
                  )

dimStart<-dim(startOUT)

biallelic<-array(
  rep(1),
  dimStart)

for(i in 1:nrow(biallelic) ) {
  for (k in 1:ncol(biallelic)){
ifelse ((startOUT[i,k,1] != startOUT[i,k,2]),biallelic[i,k,2]<-2 ,
                                             biallelic[i,k,2]<-1)
  }
}

return(biallelic)
}

```

Parameters

These are the parameters needed for the examples below

We use the base random seed function as a pseudo-random generator for reproducibility

```
set.seed(1)
```

We then build our *glads* populations using the `initial.struct` function included in the *glads* package

```

initial.population.size=400 #Initial population size
n.loci=300 #Number of loci
n.alleles.per.locus=2 #Number of alleles; biallelic for mutation

```

```

start.1 <- start.2 <- initial.struct(initial.population.size,n.loci,
n.alleles.per.locus)
start <- list(start.1, start.1) #Initial populations

```

The selected loci are identified by the following code, and the breeding value is then given

```

loci.pos=1:n.loci          # Loci position ;creates a sequence 1 to 300
add.loci=50                #Number of additive loci
fitness.pos <- 125:(125+(add.loci-1)) #Additive loci position

```

The breeding values are then given

```

bvsM = t(                  # Breeding value of additive loci
  array(
    seq(0,1, length = n.alleles.per.locus),
  )
)

```

The fitness function and demographic parameters are assigned to each population

```

sex.ratio <- 0.5
e.v=0.01          # Stochastic environmental variant
d.v=1             # Stochastic demographic variant

recom.map= c(rep(0.5, 299))
Linked<-c(60:69, 150:159, 230:239)

#Linked loci; the values between the : are in a sequence
recom.map[Linked]<-0.0001

#Recombination map with linked loci;only the values stated above take

mutationR <- 1.1*10^-4

# mutation rate # (400*600)*(1.1*10^-4) = 26.4
# ~ approx. 26 mutations per gen per pop (52.8 both pop.s/gen)

param.z1M <- list(
  sex.ratio,      #0.5
  fitness.pos,    #125:174 length=50
  bvsM,           #matrix 20vals, rep.x50
  add.loci,       #50
  e.v            #Stochastic environmental variant,0.01 # σ-env eq.1
)

param.z2M <- param.z1M

param.w1m <- list( b0 = 6,

```

```

        b1.trait1 = 0.6, b1.trait2 = 0.6,
        b2.trait1 = 1.6 , b2.trait2 = 1.6,
        b3 = 0.006, d.v = d.v, add.loci = add.loci,
        trait.correlation = 0
    )

    param.w2m <- list( b0 = 6,
        b1.trait1 = 0.6, b1.trait2 = 0.1,
        b2.trait1 = 1.6 , b2.trait2 = 1.6,
        b3 = 0.006, d.v = d.v, add.loci = add.loci,
        trait.correlation= 0
    )

    # trait 1 is under parallel selection, as b1.trait1 is the same
    # trait 2 is under divergent selection, as b1.trait1 is different

    mtrate<-0.1

    migration.rate<- 0.01

    trait1.loci <- c( 125:134, 145:154, 165:174)

    trait2.loci<- c( 135:144, 155:164)

    n.gens<-1000

```

Example 1: epistasis probability method

This simulation runs the epistasis probability method, which took 18.6 seconds in Rstudio in a MAC OS 15.1 with 64 bits 4.4 GHz

Here, a negative custom epistatic fitness function

```

h<-function(x) {
  -(2 * x)
}

```

The simulation is run using the code below

```

examplePR<-evolveM(
  x = start,                # initial genetic structure of pop.
  time = n.gens,            # number of generations
  type = "additive",        # type of simulated evolution
  recombination = "map",    # type of recombination
  recom.rate = recom.map,   # vector, no. loci - 1

```

```

migration.rate = migration.rate,
mutation.rate = mtrate,
param.z = list(param.z1M, param.z2M), #phenotype function parameters
param.w = list(param.w1m, param.w2m), # fitness function parameters
DMI.type= "probability", # method of modelling epistasis
# epistasis.map =          # for epistatic map method
pDMI = 0.5,                # DMI probability
h=h,                       # custom epistasis function
trait1.loci=trait1.loci, # loci that contribute to trait 1
trait2.loci=trait2.loci  # loci that contribute to trait 2

)

dim(examplePR[[1]])
dim(examplePR[[2]])

```

Analysis for the *glads* output

First we need to replace values in the output with 1s and 2s

```

exampleP1R <- examplePR[[1]] # sub-pop 1
exampleP2R <- examplePR[[2]] # sub-pop 2

# these are non-DMI mutations converted to 1,2 allele of equivalent
bvs

exampleP1R[ exampleP1R == 4 ] <- 1
exampleP2R[ exampleP2R == 4 ] <- 1

exampleP1R[ exampleP1R == 8 ] <- 1
exampleP2R[ exampleP2R == 8 ] <- 1

exampleP1R[ exampleP1R == 3 ] <- 2
exampleP2R[ exampleP2R == 3 ] <- 2

exampleP1R[ exampleP1R == 7 ] <- 2
exampleP2R[ exampleP2R == 7 ] <- 2

```

we then rebind the population

```
exampleP12Rm <- list(exampleP1R, exampleP2R)
```

we need to convert the data from struct to pegas

```
exampleP12R <- struct2pegas(exampleP12Rm)
```

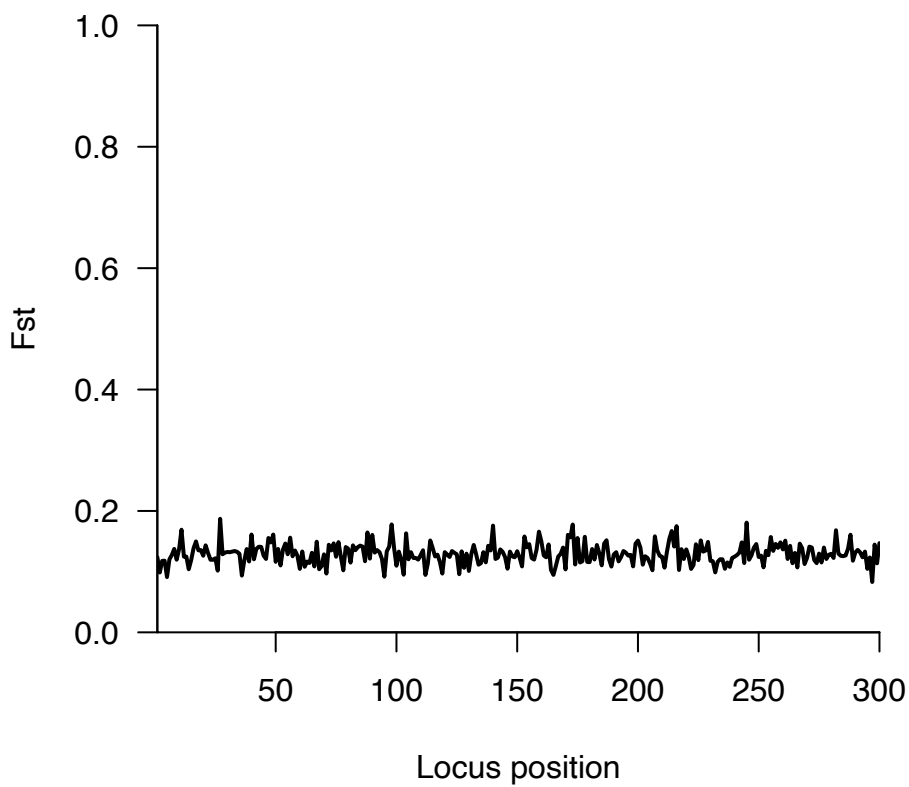
Then we need to calculate F_{ST} between the populations

```
FSTPR <- pegas::Fst(exampleP12R)
```

An example script for producing a figure of the average F_{ST} is provided below

```
plot( loci.pos,  
      FSTPR[, "Fst"],  
      pch=16,  
      bty="l",  
      xlab="Locus position",  
      main= 'Example 1',  
      ylab="Fst",  
      ylim=c(0,1),  
      type="l",  
      xaxs="i",  
      yaxs="i",  
      las=1,  
      lwd=2  
)
```

Example 1



Example 2: epistasis map method

First we need to define the epistatic map, which took 8.3 minutes in Rstudio in a MAC OS 15.1 with 64 bits 4.4 GHz

```
epiMAP<-matrix(  
  data=sample(0:1,n.loci*n.loci,replace =T),  
# matrix is of 1 and 0 ; 1 is a interaction  
  nrow=n.loci,  
  ncol=n.loci,  
  byrow=TRUE  
)
```

The code below runs the simulation

```
exampleMP<-evolveM(  
  x = start,          # initial genetic structure of pop.  
  time = n.gens,      # number of generations  
  type = "additive",  # type of simulated evolution  
  recombination = "map", # type of recombination  
  recom.rate = recom.map, # vector, no. loci - 1  
  migration.rate = migration.rate,  
  mutation.rate = mtrate,  
  param.z = list(param.z1M, param.z2M),# phenotype function parameters  
  param.w = list(param.w1m, param.w2m),# fitness function parameters  
  DMI.type= "epistasis", # method of modelling epistasis, here map  
  epistasis.map = epiMAP, # for epistatic map method  
  pDMI = 0.5,          # DMI probability  
  h=h,                # custom epistasis function  
  trait1.loci=trait1.loci, # loci that contribute to trait 1  
  trait2.loci=trait2.loci # loci that contribute to trait 2  
)
```

Analysis for the *glads* output

First, we need to replace values in the output with 1s and 2s

```
exampleM1P <- exampleMP[[1]] # sub-pop 1  
  
exampleM2P <- exampleMP[[2]] # sub-pop 2  
  
#these are non-DMI mutations converted to 1,2 allele of equivalent bvs  
  
exampleM1P[ exampleM1P == 4 ] <- 1  
exampleM2P[ exampleM2P == 4 ] <- 1  
  
exampleM1P[ exampleM1P == 8 ] <- 1  
exampleM2P[ exampleM2P == 8 ] <- 1  
  
exampleM1P[ exampleM1P == 3 ] <- 2
```

```
exampleM2P[ exampleM2P == 3 ] <- 2
exampleM1P[ exampleM1P == 7 ] <- 2
exampleM2P[ exampleM2P == 7 ] <- 2
```

we then rebind the population

```
exampleM12Pm <- list(exampleM1P, exampleM2P)
```

we need to convert the data from struct to pegas

```
exampleM12P <- struct2pegas(exampleM12Pm)
```

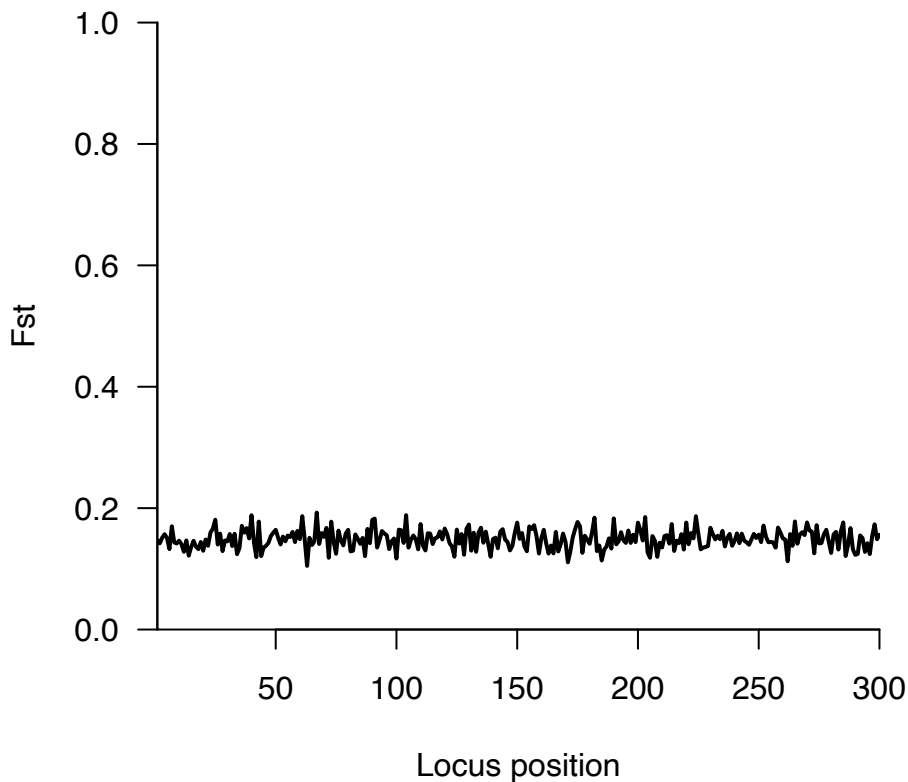
Then we need to calculate F_{ST} between the populations

```
FST <- pegas::Fst(exampleM12P)
avFst <- sum(FST)/nrow(FST) # average FST
```

An example script for producing a figure of the average F_{ST} is provided below

```
plot(loci.pos,
     FST[, "Fst"],
     pch=16,
     bty="l",
     xlab="Locus position",
     main='Example 2',
     ylab="Fst",
     ylim=c(0,1),
     type="l",
     xaxs="i",
     yaxs="i",
     las=1,
     lwd=2
)
```

Example 2



Example 3 : geno2glads

This example implements the *geno2glads* function for implementing sequence data (in the *.geno*) into the *glads* model. This sequence data is then applied to the *glads* model *evolve* function, using the DMI probability method outlined in example 1.

This example took 18.4 seconds in Rstudio in a MAC OS 15.1 with 64 bits 4.4 GHz

First, a *.geno* file is loaded into R

```
Geno <- read.table("165-129K_ZF_ified_17.geno", header=F, comment.char = "")
```

The *.geno* dataset is then formatted in R

```
Geno.t <- t(Geno[, -1])          # trim to get rid of row numbering
Geno.t[1,1] <- "ID"
colnames(Geno.t) <- Geno.t[1,]  # make col 1 colnames
```

```

Geno.t<-Geno.t[,-1,]          # then delete col 1
rownames(Geno.t) <- Geno.t[,1] # make row 1 rowname
Geno.t<-Geno.t[,,-1]         # then delete row 1 # reshaping data
loci.pos<-as.numeric(colnames(Geno.t)) # create vector of col.names
loci.pos<-loci.pos-loci.pos[1] + 1   # Loci position
chromo_mb<-max(loci.pos)            # Chromosome size
crossover <- 6.0/100000000.0

# Average recombination rate, here 6 * 10^-8 cM/MB

seq_Data<- Geno[ c(1, sample(2:nrow(Geno),300) ) , ]

# sample that keeps row 1 for structure, but samples remaining rows

```

The *geno2glads* function is then used to generate the starting populations for the *evolve* function. Here, randomly generated populations (using a pseudo-replicator seed through the *SEED* parameter) generates populations that have the same allele frequency as the data in the *.geno* file.

```

start12s<- geno2glads(seqData=seq_Data, # seqdata = .geno file
                      SEED=1,           # seed used
                      ini.ind=400)      # pop. size

```

In this example, two identical populations are applied to *glads*

```

start.1<-start12s
start.2<-start12s

```

The *evolve* function is simulated accordingly

```

exampleGG <-evolveM(
  x = list(start.1, start.2), # initial genetic structure of pop.
  time = n.gens,             # number of generations
  type = "additive",          # type of simulated evolution
  recombination = "map",      # type of recombination
  recom.rate = recom.map,     # vector, no. loci - 1
  migration.rate = migration.rate,
  mutation.rate = mtrate,
  param.z = list(param.z1M, param.z2M), # phenotype function
  param.w = list(param.w1m, param.w2m), # fitness function
  DMI.type= "probability",    # method of modelling epistasis
# epistasis.map =             # for epistatic map method
pDMI = 0.5,                  # DMI probability

```

```

h=h,                                # custom epistasis function
trait1.loci=trait1.loci, # loci that contribute to trait 1
trait2.loci=trait2.loci # loci that contribute to trait 2
)

```

Analysis for the *glads* output

First, we need to replace values in the output with 1s and 2s

```

exampleG1G <- exampleGG[[1]] # sub-pop 1

exampleG2G <- exampleGG[[2]] # sub-pop 2

exampleG1G[ exampleG1G == 4 ] <- 1
exampleG2G[ exampleG2G == 4 ] <- 1

exampleG1G[ exampleG1G == 8 ] <- 1
exampleG2G[ exampleG2G == 8 ] <- 1

exampleG1G[ exampleG1G == 3 ] <- 2
exampleG2G[ exampleG2G == 3 ] <- 2

exampleG1G[ exampleG1G == 7 ] <- 2
exampleG2G[ exampleG2G == 7 ] <- 2

```

we then rebind the population

```
exampleG12Gm <- list(exampleG1G,exampleG2G) # rebind populations
```

we need to convert the data from struct to pegas

```
exampleG12G <- struct2pegas(exampleG12Gm)
```

Then we need to calculate F_{ST} between the populations

```
FSTGG <- pegas::Fst(exampleG12G)
```

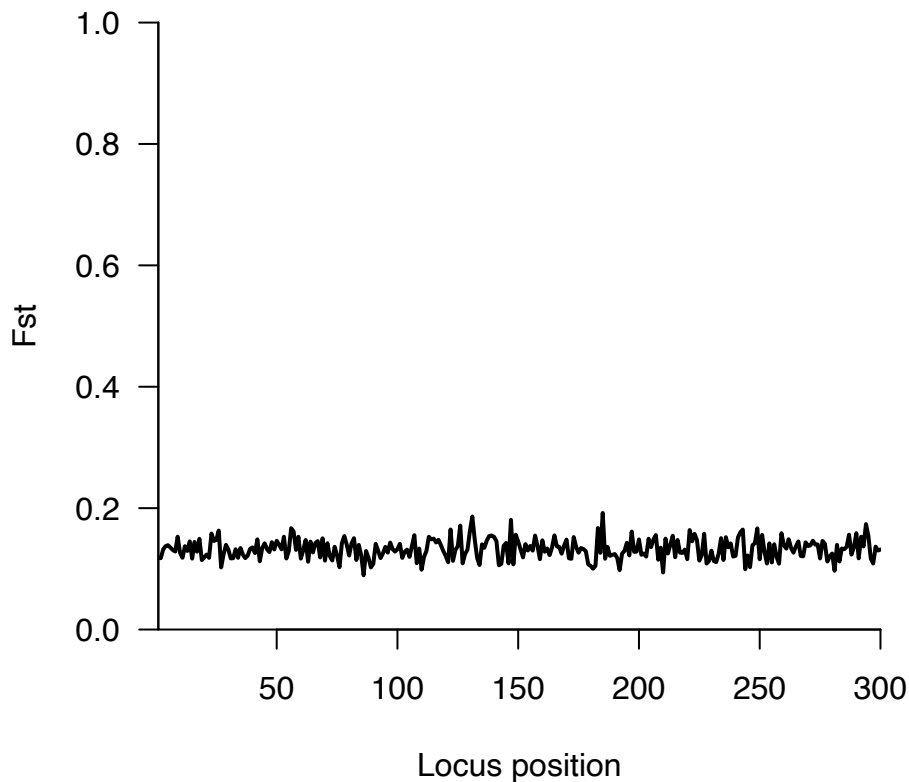
An example script for producing a figure of the average F_{ST} is provided below

```

plot( 1:n.loci ,
      FSTGG[, "Fst"],
      pch=16,
      bty="l",
      xlab="Locus position",
      main= 'Example 3',
      ylab="Fst",
      ylim=c(0,1),
      type="l",
      xaxs="i",
      yaxs="i",
      las=1,
      lwd=2
)

```

Example 3



Example 4 : positive epistasis

Here, we simulate positive epistasis using the probability method, which took 19.7 seconds in Rstudio in a MAC OS 15.1 with 64 bits 4.4 GHz

First, we define a positive monotonic custom epistasis function

```
h<-function(x) {  
  (0.3 * x)  
}
```

The evolve function is simulated accordingly

```
exampleE<-evolveM(  
  x = start,           # initial genetic structure of pop.  
  time = n.gens,       # number of generations
```

```

type = "additive",      # type of simulated evolution
recombination = "map",  # type of recombination
recom.rate = recom.map, # vector, no. loci - 1
migration.rate = migration.rate,
mutation.rate = mtrate,
param.z = list(param.z1M, param.z2M),# phenotype function parameters
param.w = list(param.w1m, param.w2m),# fitness function parameters
DMI.type= "probability", # method of modelling epistasis
# epistasis.map =         # for epistatic map method
pDMI = 0.5,              # DMI probability
h=h,                     # custom epistasis function
trait1.loci=trait1.loci, # loci that contribute to trait 1
trait2.loci=trait2.loci  # loci that contribute to trait 2

)

```

Analysis for the *glads* output

First, we need to replace values in the output with 1s and 2s

```

example1E <- exampleE[[1]] # sub-pop 1

example2E <- exampleE[[2]] # sub-pop 2

#these are non-DMI mutations converted to 1,2 allele of equivalent bvs

example1E[ example1E == 4 ] <- 1
example2E[ example2E == 4 ] <- 1

example1E[ example1E == 8 ] <- 1
example2E[ example2E == 8 ] <- 1

example1E[ example1E == 3 ] <- 2
example2E[ example2E == 3 ] <- 2

example1E[ example1E == 7 ] <- 2
example2E[ example2E == 7 ] <- 2

```

we then rebind the population

```
example12EE <- list(example1E, example2E)
```

we need to convert the data from struct to pegas

```
exampleP12E <- struct2pegas(example12EE)
```

Then we need to calculate F_{ST} between the populations

```
FSTPR <- pegas::Fst(exampleP12E)
```

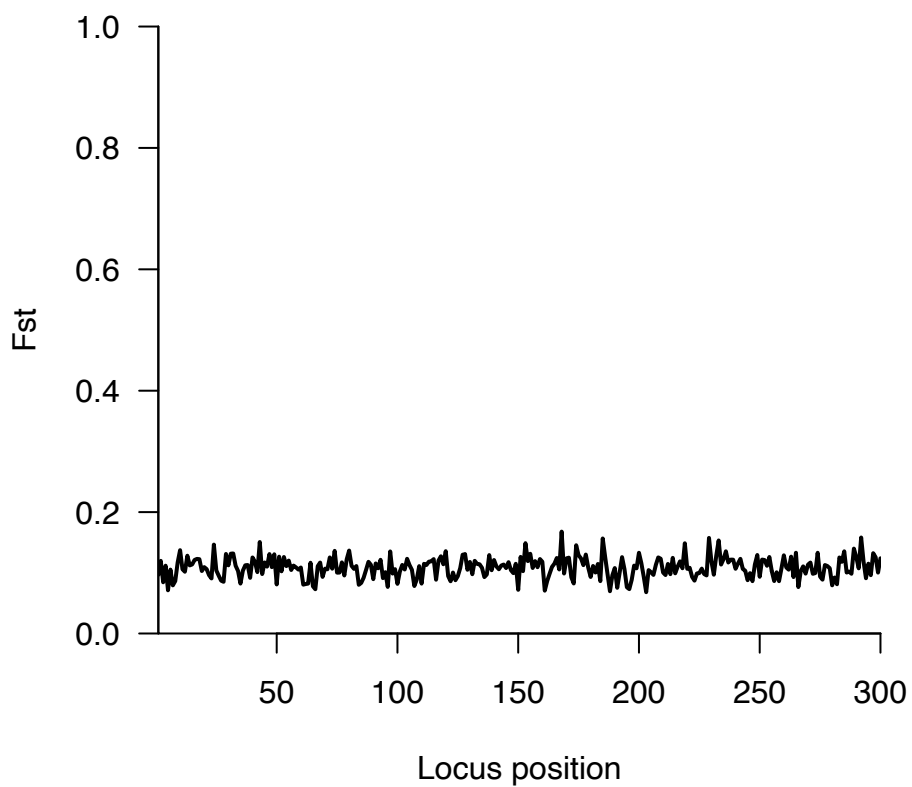
An example script for producing a figure of the average F_{ST} is provided below

```

plot( loci.pos,
      FSTPR[, "Fst"],
      pch=16,
      bty="l",
      xlab="Locus position",
      main= 'Example 4',
      ylab="Fst",
      ylim=c(0,1),
      type="l",
      xaxs="i",
      yaxs="i",
      las=1,
      lwd=2
    )

```

Example 4



C. Supplementary material: Chapter 4

Appendix 1

Supporting Information Appendix S1 for the manuscript:

Effects of selection and hybridisation barriers on the genomic landscape of introgression

Questions to: Hisham A.A. Ali. Department of Zoology, University of Oxford.

Email: hisham.ali@biology.ox.ac.uk

This vignette provides a description of how to set up *glads* in R and four examples implementing the modifications from this manuscript.

Example 1 : Selection

Example 2 : Reduced hybrid fitness

Example 3 : Heterosis

Example 4 : Genes of large effect

R set up

Within R, the following packages need to be installed:

```
library('ape')
library('ade4')
library('adegenet')
library('pegas')
library('Rcpp')
library('abind')
library('parallel')
library('progress')
library('ggplot2')
library('dplyr')
library('caTools')
```

The *glads* package can be cloned from Github in terminal by running the following commands

```
git clone https://github.com/eriqande/glads
```

```
R CMD INSTALL --no-multiarch --with-keep.source glads
```

After installing *glads*, the library can be loaded into R

```
library('glads')
```

New R functions

The modified `migrate` function

```
migrate_uniDIR <-
function (G1, G2, rate1to2, rate2to1)
{
  n1 <- dim(G1)[1]
  n2 <- dim(G2)[1]
  L <- dim(G1)[2]
  g <- 2
```

```

p1 <- ifelse(runif(n1) < rate1to2, 2, 1)
p2 <- ifelse(runif(n2) < 0, 1, 2)
ret <- rcpp_dispersal_placement(G1, G2, dim(G1), dim(G2), p1, p2)
dim(ret[[1]]) <- c(sum(c(p1, p2) == 1), L, g)
dim(ret[[2]]) <- c(sum(c(p1, p2) == 2), L, g)
ret
}

```

The modified `newborns` function

```

newborns_HR <-
  function (x, recombination, type,
            htz.locus, hm.strength, ht.strength, ht.sex # Haldane's rule
            )
  {
    struct          <- x[[1]]
    recom.rate      <- x[[2]]
    init.sex        <- x[[3]]
    mutation.rate   <- x[[4]]
    loci.pos        <- x[[5]]
    chromo_mb       <- x[[6]]
    param.z         <- x[[7]]
    param.w         <- x[[8]]
    fun             <- x[[9]]
    pop             <- x[[10]]
    time.t          <- x[[11]]
    switch(type, constant = {
mean.fitness <- 2
n <- nrow(struct)
if (n%%2 == 0) {
  z <- as.data.frame(cbind(sex = sample(rep(0:1, each = n/2)) + 1))
  rownames(z) = rownames(struct)
} else {
  neven = floor(n) + floor(n)%2 - 2
  z <- as.data.frame(cbind(sex = c(sample(rep(0:1, each = neven/2)),
  sample(0:1, 1)) + 1))
  rownames(z) = rownames(struct)
}

if (!is.null(init.sex)) {
  z <- as.data.frame(cbind(sex = init.sex))
  rownames(z) = rownames(struct)
}
z$z <- rep(mean.fitness, n)
},
dynamic = {
  sex.ratio <- param.z[[1]]
  mean.fitness <- param.w[[1]]

```

```

d.d <- param.w[[2]]
n <- nrow(struct)
z <- as.data.frame(cbind(sex = rbinom(n, 1, sex.ratio) + 1))
rownames(z) = rownames(struct)
z$z <- round(pmax(rpois(n, lambda = mean.fitness) - d.d * n, 0))
    },
additive = {
  sex.ratio <- param.z[[1]]
  fitness.pos <- param.z[[2]]
  bvs <- param.z[[3]]
  add.loci <- param.z[[4]]
  e.v <- param.z[[5]]
  b0 <- param.w[[1]]
  b1 <- param.w[[2]]
  b2 <- param.w[[3]]
  b3 <- param.w[[4]]
  d.v <- param.w[[5]]
  n.loci <- param.w[[6]] # fitness change
  struct.fit <- struct[, fitness.pos, ]
  zs <- phenotype(struct.fit, bvs, add.loci, sex.ratio, e.v)
  fit <- fitnessAdj(zs, dim(zs)[1], b0, b1, b2, b3, d.v, n.loci, time.t)
# fitness change #change fun name &add n.loci

  if ( length(fit) >=1 ) {

    for (h in 1: length(fit) ) {

if( struct[h,htz.locus,1] != struct[h,htz.locus,2] ) { # heterozygote

  if( zs[, "sex"][h] == ht.sex ) { # heterogametic sex

    fit[h] <- round( ( fit[h] * ht.strength ),0 )

    fit[h] <- ifelse( fit[h] < 0, 0, fit[h])

  }

  else if ( zs[, "sex"][h] != ht.sex ) { # homogametic sex

    fit[h] <- round( ( fit[h] * hm.strength ),0 )

    fit[h] <- ifelse( fit[h] < 0, 0, fit[h])

  }

}

}

}

}

z <- cbind(sex = zs[, "sex"], z = fit)

```

```

    },
    custom = {
  zs <- do.call(fun[1], append(list(struct = struct), param.z))
  fit <- do.call(fun[2], append(list(time.t = time.t, z = zs[, "z"],
                                sex = zs[, "sex"],
                                n = dim(zs)[1]), param.w) )
  z <- cbind(sex = zs[, "sex"], z = fit)
    },

  stop("Invalid type of evolution. Current options are 'constant', 'dynamic', 'additive', and 'custom'"))

  if (length(unique(z[, "sex"])) > 1 && !0 %in% aggregate(z ~ sex, data
= z, sum)$z) {
    pairs <- mating(z, struct)
    mums <- pairs[[1]]
    dads <- pairs[[2]]

    if (!is.null(mutation.rate)) {
      mums <- mutation(mums, mutation.rate)
      dads <- mutation(dads, mutation.rate)
    }
    struct.rt <- array(NA, dim(mums))
    switch(recombination, map = {
      struct.rt[, , 1] <- rcpp_recombo_segagate(mums,
                                                dim(mums), recom.rate)
      struct.rt[, , 2] <- rcpp_recombo_segagate(dads,
                                                dim(dads), recom.rate)
    },
    average = {

      struct.rt[, , 1] <- rcpp_recombo_segagate_expo(mums,
                                                      dim(mums), loci.pos, chromo_mb, recom.rate)
      struct.rt[, , 2] <- rcpp_recombo_segagate_expo(dads,
                                                      dim(dads), loci.pos, chromo_mb, recom.rate)
    },

    stop("Invalid recombination type. Current options are 'map' or 'average'"))
  }

  else {
    struct.rt <- struct[0, , ]
  }

  return(struct.rt)
}

```

The modified `evolve` function

```
evolve_mod <-  
function (  
  x, time, type = c("constant", "dynamic", "additive","custom"),  
  recombination = c("map", "average"), recom.rate, loci.pos = NULL,  
  chromo_mb = NULL, init.sex = NULL, migration.rate = NULL,  
  mutation.rate = NULL, param.z = NULL, param.w = NULL,  
  fun = c(phenotype = NULL, fitness = NULL),  
  admixture = list(taxa = NULL, dd = FALSE, nmin = 10),  
  htz.locus,hm.strength, ht.strength,ht.sex          # Haldane's rule  
  )  
{  
  npop    <- length(x)  
  struct <- x  
  
  if (is.na(time[2]) == F) {  
    each <- time[2]  
  }  
  else {  
    each <- time[1]  
  }  
  time <- time[1]  
  if (is.list(x) == F)  
    stop("Initial populations should be a list of arrays.", call. = F)  
  
  if (recombination == "average" && (is.null(loci.pos)  
    || is.null(chromo_mb)))  
    stop("Loci position or size of the chromosome is missing with  
no default for 'average' recombination.", call. = F)  
  
  if ((recombination == "map" && length(recom.rate) != (dim(x)[[1]])[2] -
```

```

1)) || (recombination == "average" && length(recom.rate) != 1))

stop("Incorrect recombination rate for the type of recombination ('map
' or 'average').", call. = F)

if (type == "custom" && (is.character(fun) == FALSE || length(fun) !=
2))
stop("One or both custom function names are missing or 'fun' is not we
ll defined. 'fun' is a character vector of length 2, with the name of
the phenotype and fitness functions e.g. c('phenotype', 'fitness').",
call. = F)

if (!is.null(mutation.rate)) {
  if (sum(sapply(1:npop, function(i) {

    length(table(x[[i]])) != 2})) != 0) {
    mutation.rate = NULL

warning("Mutation rate is ignored for non-biallelic genetic structures
. The genetic structure input should contains integers of values 1 or
2.", call. = F)
    }

  else if (sum(sapply(1:npop, function(i) {
    1 %in% x[[i]] && 2 %in% x[[i]]
  }) == FALSE) != 0) {
    mutation.rate = NULL

warning("Mutation rate is ignored for non-biallelic genetic structures
. The genetic structure input should contains integers of values 1 or
2.", call. = F)
    }
  }

  if (npop == 1 && !is.null(migration.rate)) {
    migration.rate = NULL

warning("Migration is ignored for simulations with a single population
.", call. = F)
  }

  if (!is.null(migration.rate)) {
    dispchange <- time
    if (is.list(migration.rate))

```

```

        dispchange <- c(migration.rate[[1]], dispchange)
        disp <- mmatrix(migration.rate, npop)
      }

      if (!is.null(init.sex) && (length(init.sex) != npop
        || sum(sapply(1:length(x),
function(i) {
  nrow(x[[i]]) != length(init.sex[[i]])  }))) != 0))

stop("There is not enough females or males assigned to the initial pop
ulations.", call. = F)

pb <- progress_bar$new(format = " Work in progress [:bar] :percent eta
: :eta", total = time, clear = FALSE, width = 100)

switch(recombination, map = {
  res <- list()
  if (each != time) res[["0"]] <- struct
  for (i in 1:time) {
    if (i > 1) {
      init.sex <- NULL
    }
  }

y <- lapply(1:npop, function(j) { append(list(struct[[j]]),
  list(recom.rate, init.sex[[j]],
    mutation.rate, loci.pos = NULL,

    chromo_mb = NULL, param.z[[j]], param.w[[j]], fun, j, i
  )),  })

out <- lapply(y, newborns_HR,

  recombination = recombination, type = type,
  htz.locus, hm.strength, ht.strength, ht.sex

)
if (!is.null(migration.rate)) {
  move <- combn(1:npop, 2)
  p <- which(i <= dispchange)[[1]]
  dispp <- disp[, , p]
  if (!is.null(admixture$taxa)) {
    dispp <- admix(x = out, mr = dispp, taxa = admixture$taxa,
      dd = admixture$dd, nmin = admixture$nmin)

```

```

    }

    for (j in 1:ncol(move)) {
      n1 <- move[1, j]
      n2 <- move[2, j]
      rate1to2 <- dispp[n1, n2]
      rate2to1 <- dispp[n2, n1]
      outd <- migrate_uniDIR(out[[n1]], out[[n2]], rate1to2, rate2to1)
      out[[n1]] <- outd[[1]]
      out[[n2]] <- outd[[2]]
    }
  }

  struct <- out
  if (i%%each == 0) {
    if (each != time) {
      class(struct) <- c("struct", "array")
      res[[paste(i)]] <- struct
    } else {
      class(struct) <- c("struct", "array")
      res <- struct
    }
  }
  pb$tick()
}

}, average = {
  res <- list()
  if (each != time) res[["0"]] <- struct
  for (i in 1:time) {
    if (i > 1) {
      init.sex <- NULL
    }
  }
}

y <- lapply(1:npop, function(j) {append(list(struct[[j]]),
  list(recom.rate, init.sex[[j]], mutation.rate, loci.pos,
    chromo_mb, param.z[[j]], param.w[[j]], fun, j, i))
}))

out <- lapply(y, newborns_HR,
  recombination = recombination, type = type,
  htz.locus, hm.strength, ht.strength, ht.sex )

if (!is.null(migration.rate)) {
  move <- combn(1:npop, 2)
  p <- which(i <= dispchange)[[1]]
  dispp <- disp[, , p]
  if (!is.null(admixture$taxa)) {

```

```

        dispp <- admix(x = out, mr = dispp, taxa = admixture$taxa,
                      dd = admixture$dd, nmin = admixture$nmin)
      }
    for (j in 1:ncol(move)) {
      n1 <- move[1, j]
      n2 <- move[2, j]
      rate1to2 <- dispp[n1, n2]
      rate2to1 <- dispp[n2, n1]
      outd <- migrate_uniDIR(out[[n1]], out[[n2]], rate1to2, rate2to1
                             )
      out[[n1]] <- outd[[1]]
      out[[n2]] <- outd[[2]]
    }
  }
  struct <- out
  if (i%%each == 0) {
    if (each != time) {
      class(struct) <- c("struct", "array")
      res[[paste(i)]] <- struct
    } else {
      class(struct) <- c("struct", "array")
      res <- struct
    }
  }
  pb$tick()
},
),
stop("Invalid recombination type. Current options are 'map' or 'average'"))
  return(res)
}

```

The modified `fitness` function

```

fitnessAdj <-
function(z, N, b0, b1, b2, b3, d.v, add.loci, time.t)
{
  ze <- z[, "z"]
  ng <- add.loci          # fitness change # use ng
  a <- b0
  b <- b1
  c <- b2

  w <- round( a*exp(-((ze-b*ng)^2)/(2*(c*ng)^2)) ) - b3*N +
                                                     rnorm(N, 0, d.v) , 0)

  w <- ifelse(w<0, 0, w)
  return(w)
}

```

Parameters

These are the parameters needed for the examples below

We use the base random seed function as a pseudo-random generator for reproducibility

```
set.seed(1)
```

We then build our *glads* populations using the `initial.struct` function included in the *glads* package

```
initial.population.size=400 #Initial population size
n.loci=300                  #Number of loci
n.alleles.per.locus=2      #Number of alleles; biallelic for mutation

start.1 <- start.2 <- initial.struct(initial.population.size,n.loci,
n.alleles.per.locus)
start <- list(start.1, start.1) #Initial populations
```

The selected loci are identified by the following code, and the breeding value is then given

```
loci.pos=1:n.loci          # Loci position
add.loci=50                # Number of additive loci
fitness.pos <- 125:(125+(add.loci-1)) # Additive loci position
```

The linkage map is given by

```
recom.map <- c( rep(0.1,149), # linkage pattern, for 300 loci, linked
               rep(0.5,150)  # unlinked
             )
```

The fitness function parameters are then defined for the populations

```
b1 <- 0.4

d.v <- 1                      #Stochastic demographic variant

param.w1 <- list(b0 = 4, b1 = b1, b2 = 0.8, b3 = 0.005, d.v = d.v, add.
loci = 200)

param.w2 <- list(b0 = 4, b1 = b1, b2 = 0.8, b3 = 0.004, d.v = d.v, add.
loci = 200)
```

Breeding values are allocated for the alleles

```

bvs.ad <- array( rep(0.4), dim = c( add.loci, 2 ) )

bvM <- b1/2      # optimal breeding value

devitn <- 0.4    # selection (difference in fitness of breeding values
)

bvs.ad[,1] <- rnorm( add.loci,
                     mean = bvM ,
                     sd = 0.15
)

bvs.ad[,2] <- rnorm( add.loci,
                     mean = bvM + devitn ,
                     sd = 0.15
)

```

The phenotype function parameters are then defined for the populations

```

sex.ratio <- 0.5
e.v       <- 0.01 # Stochastic environmental variant

param.z1 <- list(
  sex.ratio,
  fitness.pos,
  bvs.ad,      # matrix, add.loci * number of alleles
  add.loci,
  e.v          #  $\sigma$ -env eq.1
)

param.z2 <- param.z1

```

Example 1 : Selection

This example looks at the landscape of introgression due to selection, which took 8.1 seconds in Rstudio in a MAC OS 15.1 with 64 bits 4.4 GHz

```

n.gens <- 1000
n.gens.out <- 100

test1 <- evolve_mod (
  x = start,                # initial genetic structure of pop.
  time = c(n.gens, n.gens.out), # number of gen.s; how often recorded
  type = "additive",         # type of simulated evolution
  recombination = "map",     # type of recombination
  recom.rate = recom.map,

```

```

migration.rate= 0.0005 ,      # unidir mig. pop. 1 -> 2 only
param.z = list(param.z1, param.z2), # phenotype function parameters
param.w = list(param.w1, param.w2) # fitness function parameters
,htz.locus = 250,hm.strength = 1 , ht.strength = 1, ht.sex = 1
)

```

The output of the *glads* model is then analysed accordingly to generate the genomic landscape of introgression

```

migrant_fq <-      matrix( rep(0), nrow = 2 , ncol = n.loci )
host_fq <-         matrix( rep(0), nrow = 2 , ncol = n.loci )

landscape_introg <- matrix( rep(-1), nrow = 2 , ncol = n.loci )
for (i in 1: n.loci ) {

  migrant_fq[1,i]      <- sum ( test1[["500"]][[2]] [,i,] == '1')
  host_fq[1,i]         <- sum ( test1[["500"]][[2]] [,i,] == '2')
  landscape_introg[1,i] <- migrant_fq[1,i] / ( migrant_fq[1,i] + host
_fq[1,i] )

  migrant_fq[2,i]      <- sum ( test1[["1000"]][[2]] [,i,] == '1')
  host_fq[2,i]         <- sum ( test1[["1000"]][[2]] [,i,] == '2')
  landscape_introg[2,i] <- migrant_fq[2,i] / ( migrant_fq[2,i] + host
_fq[2,i] )

}

```

An example script for producing a figure of the genomic landscape of introgression is provided below

```

plot( loci.pos,
      landscape_introg[2,],
      pch=16,
      bty="l",
      xlab="Locus position",
      ylab="Introgression Proportion",
      main= 'Example 1',
      ylim=c(0,1),
      type="l",
      xaxs="i",
      yaxs="i",
      las=1,
      lwd=2
)

```

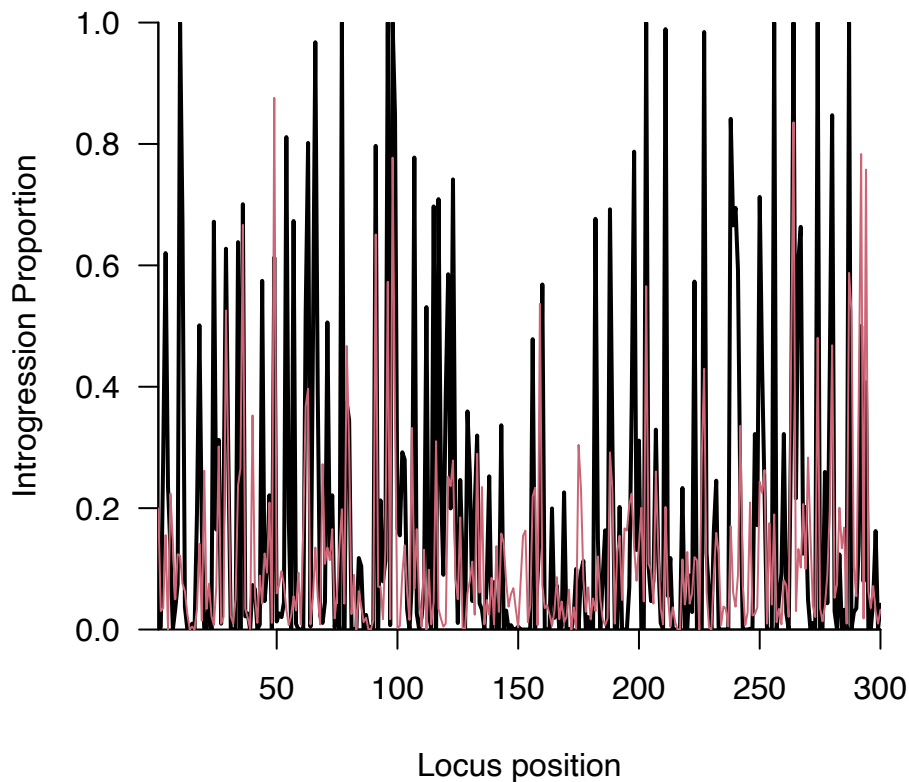
The landscape of introgression for the other population is added by the pink line

```

lines ( loci.pos, landscape_introg[1,],type='l', col='#CC6677' )

```

Example 1



Example 2: Reduced hybrid fitness

In this example, only hybrids of the homogametic sex have a lower fitness in hybrids

In this example, the male is the heterogametic sex (individuals assigned = 1)

This example took 7.6 seconds in Rstudio in a MAC OS 15.1 with 64 bits 4.4 GHz

```
n.gens <- 1000
n.gens.out <- 100

test2 <- evolve_mod (
  x = start,                                # initial genetic structure of pop.
  time = c(n.gens, n.gens.out),             # number of gen.s; how often recorded
  type = "additive",                        # type of simulated evolution
```

```

recombination = "map",      # type of recombination
recom.rate = recom.map,
migration.rate= 0.0005 ,    # unidir mig. pop. 1 -> 2 only
param.z = list(param.z1, param.z2), # phenotype function parameters
param.w = list(param.w1, param.w2), # fitness function parameters
htz.locus = 250,hm.strength = 1 , ht.strength = 0.4, ht.sex = 1
)

```

The output of the *glads* model is then analysed accordingly to generate the genomic landscape of introgression

```

migrant_2fq <-      matrix( rep(0), nrow =  2 , ncol = n.loci  )
host_2fq <-         matrix( rep(0), nrow =  2 , ncol = n.loci  )
landscape_2introg <- matrix( rep(-1), nrow =  2 , ncol = n.loci  )

for (i in 1: n.loci  ) {

  migrant_2fq[1,i]      <- sum ( test2[["500"]][[2]] [,i,] == '1')
  host_2fq[1,i]         <- sum ( test2[["500"]][[2]] [,i,] == '2')
  landscape_2introg[1,i] <- migrant_2fq[1,i] / ( migrant_2fq[1,i] + h
ost_2fq[1,i]          )

  migrant_2fq[2,i]      <- sum ( test2[["1000"]][[2]] [,i,] == '1')
  host_2fq[2,i]         <- sum ( test2[["1000"]][[2]] [,i,] == '2')
  landscape_2introg[2,i] <- migrant_2fq[2,i] / ( migrant_2fq[2,i] + h
ost_2fq[2,i]          )

}

```

An example script for producing a figure of the genomic landscape of introgression is provided below

```

plot( loci.pos,
      landscape_2introg[2,],
      pch=16,
      bty="l",
      xlab="Locus position",
      ylab="Introgression Proportion",
      main= 'Example 2',
      ylim=c(0,1),
      type="l",
      xaxs="i",
      yaxs="i",
      las=1,
      lwd=2
)

```

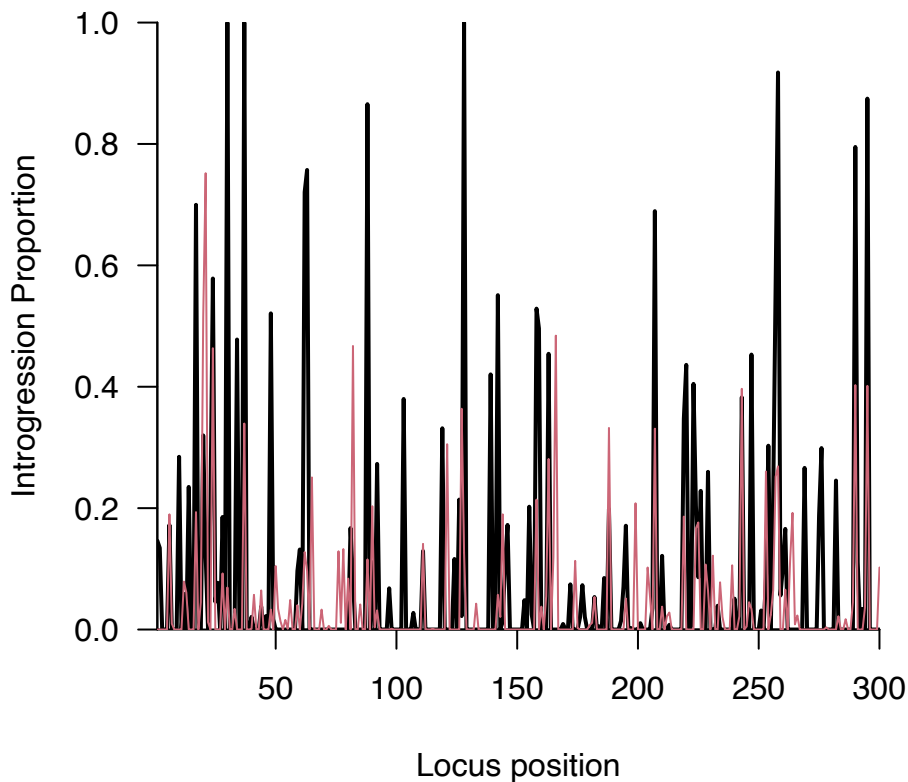
The landscape of introgression for the other population is added by the pink line

```

lines ( loci.pos, landscape_2introg[1,],type='l', col='#CC6677' )

```

Example 2



Example 3 : Heterosis

in this model, the hybrids of the heterogametic sex display heterosis (hybrid vigour)

in this model, female is the heterogametic sex

This example took 9.6 seconds in Rstudio in a MAC OS 15.1 with 64 bits 4.4 GHz

```
n.gens <- 1000
n.gens.out <- 100

test3 <- evolve_mod (
  x = start,                                # initial genetic structure of pop.
  time = c(n.gens, n.gens.out)              # number of gens.; how often recorded
  type = "additive",                        # type of simulated evolution
  recombination = "map",                    # type of recombination
  recom.rate = recom.map,
  migration.rate = 0.0005,                  # unidir mig. pop. 1 -> 2 only
  param.z = list(param.z1, param.z2),      # phenotype function parameters
  param.w = list(param.w1, param.w2)       # fitness function parameters
  , htz.locus = 250, hm.strength = 1.3, ht.strength = 1, ht.sex = 2
)
```

The output of the *glads* model is then analysed accordingly to generate the genomic landscape of introgression

```
migrant_3fq <-      matrix( rep(0), nrow =  2 , ncol = n.loci  )
host_3fq <-        matrix( rep(0), nrow =  2 , ncol = n.loci  )
landscape_3introq <- matrix( rep(-1), nrow =  2 , ncol = n.loci  )

for (i in 1: n.loci  ) {

  migrant_3fq[1,i]      <- sum ( test3[["500"]][[2]] [,i,] == '1')
  host_3fq[1,i]         <- sum ( test3[["500"]][[2]] [,i,] == '2')
  landscape_3introq[1,i] <- migrant_3fq[1,i] / ( migrant_3fq[1,i] + h
ost_3fq[1,i] )

  migrant_3fq[2,i]      <- sum ( test3[["1000"]][[2]] [,i,] == '1')
  host_3fq[2,i]         <- sum ( test3[["1000"]][[2]] [,i,] == '2')
  landscape_3introq[2,i] <- migrant_3fq[2,i] / ( migrant_3fq[2,i] + h
ost_3fq[2,i] )

}
```

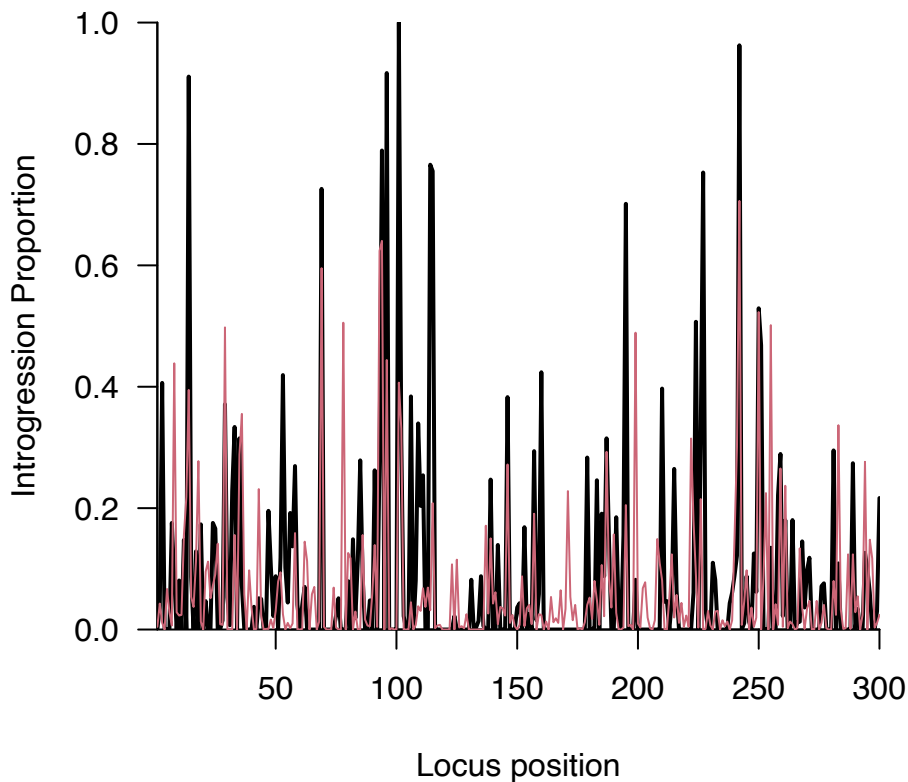
An example script for producing a figure of the genomic landscape of introgression is provided below

```
plot( loci.pos,
      landscape_3introq[2,],
      pch=16,
      bty="l",
      xlab="Locus position",
      ylab="Introgression Proportion",
      main= 'Example 3',
      ylim=c(0,1),
      type="l",
      xaxs="i",
      yaxs="i",
      las=1,
      lwd=2
    )
```

The landscape of introgression for the other population is added by the pink line

```
lines ( loci.pos, landscape_3introq[1,],type='l', col='#CC6677' )
```

Example 3



Example 4 : Genes of large effect

In this example, the breeding values are adjusted to simulate genes of large effect

Below, the breeding values are adjusted to simulate genes of large effect. In this example, 10 selected loci are simulated as large effect loci explaining 55% of the phenotype

This example took 7.4 seconds in Rstudio in a MAC OS 15.1 with 64 bits 4.4 GHz

```
no.lрге.еff <- 10 # number of large effect loci
not.lрге.еff <- length(fitness.pos) - no.lрге.еff

# number of non- large effect loci

large.effecт <- 0.55 / no.lрге.еff

# 10 loci explain 55 % of phenotype

non.large.effecт <- 0.45 / not.lрге.еff
```

```

# 40 loci explain 45 % of phenotype

large.effect.pos <- sample(length(fitness.pos),no.lрге.eff,replace = F)

weight <- rep(non.large.effect,length(fitness.pos))
weight [large.effect.pos] <- large.effect # create a vector of cand. a
nd non-cand. Weights, sum(weight) = 1

bvs.adj <- bvs.ad # create new matrix for weighted bvs

bvs.adjs <- bvs.ad * add.loci # as we are multiplying the vector by
weights, standerdise with loci number
# to correct for this - if all bvs same then same total after weightin
g

for ( i in 1:nrow(bvs.adjs) ) {

  bvs.adj[i,] <- bvs.adjs[i,] * weight[i]

}

```

Parameters for the population phenotype function are given below

```

param.z1a <- list(
  sex.ratio,      # 0.5
  fitness.pos,    # 125:174 length=50
  bvs.adj,        # matrix 20vals, rep.x50
  add.loci,       # 50
  e.v             # Stochastic environmental variant,0.01 #  $\sigma$ -env eq.1
)

param.z2a <- param.z1a

```

The model is simulated using the code below

```

n.gens      <- 1000
n.gens.out  <- 100

test4 <- evolve_mod (
  x = start,                # initial genetic structure of pop.
  time = c(n.gens,n.gens.out), # number of gen.s; how often recorded
  type = "additive",        # type of simulated evolution
  recombination = "map",    # type of recombination
  recom.rate = recom.map,
  migration.rate= 0.0005 ,  # unidir mig. pop. 1 -> 2 only
  param.z = list(param.z1a, param.z2a),# phenotype function parameters
  param.w = list(param.w1, param.w2)  # fitness function parameters
  ,htz.locus = 250,hm.strength = 1 , ht.strength = 1, ht.sex = 2
)

```

The output of the *glads* model is then analysed accordingly to generate the genomic landscape of introgression

```
migrant_4fq      <- matrix( rep(0), nrow = 2 , ncol = n.loci )
host_4fq         <- matrix( rep(0), nrow = 2 , ncol = n.loci )
landscape_4intro <- matrix( rep(-1), nrow = 2 , ncol = n.loci )

for (i in 1: n.loci ) {

  migrant_4fq[1,i] <- sum ( test4[["500"]][[2]] [,i,] == '1')
  host_4fq[1,i]    <- sum ( test4[["500"]][[2]] [,i,] == '2')
  landscape_4intro[1,i] <- migrant_4fq[1,i] / ( migrant_4fq[1,i] + h
ost_4fq[1,i] )

  migrant_4fq[2,i] <- sum ( test4[["1000"]][[2]] [,i,] == '1')
  host_4fq[2,i]    <- sum ( test4[["1000"]][[2]] [,i,] == '2')
  landscape_4intro[2,i] <- migrant_4fq[2,i] / ( migrant_4fq[2,i] + h
ost_4fq[2,i] )

}
```

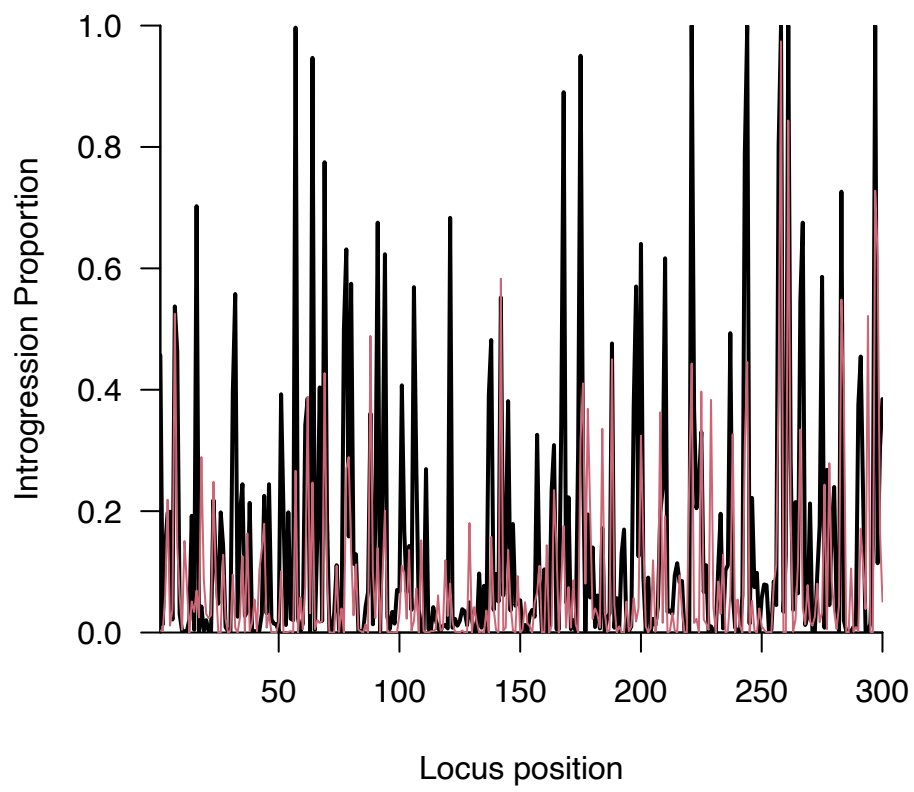
An example script for producing a figure of the genomic landscape of introgression is provided below

```
plot( loci.pos,
      landscape_4intro[2,],
      pch=16,
      bty="l",
      xlab="Locus position",
      ylab="Introgression Proportion",
      main= 'Example 4',
      ylim=c(0,1),
      type="l",
      xaxs="i",
      yaxs="i",
      las=1,
      lwd=2
    )
```

The landscape of introgression for the other population is added by the pink line

```
lines ( loci.pos, landscape_4intro[1,],type='l', col='#CC6677' )
```

Example 4



D. Supplementary material: Genomic

landscapes of divergence among island bird populations: Evidence of parallel adaptation but at different loci?

Claudia A Martin, Eleanor C Sheppard, Hisham AA Ali, Juan Carlos Illera, Alexander Suh, Lewis G Spurgin, David S Richardson

Published as:

Martin, C. A., Sheppard, E. C., **Ali, H. A.**, Illera, J. C., Suh, A., Spurgin, L. G., & Richardson, D. S. (2024). Genomic landscapes of divergence among island bird populations: Evidence of parallel adaptation but at different loci?. *Molecular Ecology*, 33(12), e17365.

ORIGINAL ARTICLE

Genomic landscapes of divergence among island bird populations: Evidence of parallel adaptation but at different loci?

Claudia A. Martin^{1,2,3}  | Eleanor C. Sheppard¹  | Hisham A. A. Ali⁴  |
Juan Carlos Illera⁵ | Alexander Suh^{1,6}  | Lewis G. Spurgin¹  | David S. Richardson¹ 

¹School of Biological Sciences, University of East Anglia, Norfolk, UK

²Terrestrial Ecology Unit, Biology Department, Ghent University, Ghent, Belgium

³School of Biological Sciences, The University of Edinburgh, Edinburgh, UK

⁴Department of Biology, Edward Grey Institute of Field Ornithology, University of Oxford, Oxford, UK

⁵Biodiversity Research Institute (CSIC-Oviedo University-Principality of Asturias), University of Oviedo, Mieres, Asturias, Spain

⁶Department of Organismal Biology – Systematic Biology, Evolutionary Biology Centre (EBC), Science for Life Laboratory, Uppsala University, Uppsala, Sweden

Correspondence

Claudia A. Martin and David S. Richardson, School of Biological Sciences, University of East Anglia, Norwich Research Park, Norfolk, UK.

Email: cmarti3@ed.ac.uk and david.richardson@uea.ac.uk

Funding information

Norwich Research Park Science Links Seed Fund; Regional Government of Asturias, Grant/Award Number: AYUD/2021/51261; Natural Environment Research Council, Grant/Award Number: NE/L002582/1 and NE/S007334/1; Spanish Ministry of Science, Innovation and Universities; European Regional Development Fund, Grant/Award Number: PGC2018-097575-B-I00

Handling Editor: David Coltman

Abstract

When populations colonise new environments, they may be exposed to novel selection pressures but also suffer from extensive genetic drift due to founder effects, small population sizes and limited interpopulation gene flow. Genomic approaches enable us to study how these factors drive divergence, and disentangle neutral effects from differentiation at specific loci due to selection. Here, we investigate patterns of genetic diversity and divergence using whole-genome resequencing (>22× coverage) in Berthelot's pipit (*Anthus berthelotii*), a passerine endemic to the islands of three north Atlantic archipelagos. Strong environmental gradients, including in pathogen pressure, across populations in the species range, make it an excellent system in which to explore traits important in adaptation and/or incipient speciation. First, we quantify how genomic divergence accumulates across the speciation continuum, that is, among Berthelot's pipit populations, between sub species across archipelagos, and between Berthelot's pipit and its mainland ancestor, the tawny pipit (*Anthus campestris*). Across these colonisation timeframes (2.1 million–ca. 8000 years ago), we identify highly differentiated loci within genomic islands of divergence and conclude that the observed distributions align with expectations for non-neutral divergence. Characteristic signatures of selection are identified in loci associated with craniofacial/bone and eye development, metabolism and immune response between population comparisons. Interestingly, we find limited evidence for repeated divergence of the same loci across the colonisation range but do identify different loci putatively associated with the same biological traits in different populations, likely due to parallel adaptation. Incipient speciation across these island populations, in which founder effects and selective pressures are strong, may therefore be repeatedly associated with morphology, metabolism and immune defence.

KEYWORDS

birds, craniofacial evolution, divergence landscape, genomic islands, immune defence, parallel adaptation, speciation

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *Molecular Ecology* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

Genetic differentiation among populations accumulates over time due to a combination of differing adaptive and neutral processes (Feder et al., 2013; Seehausen et al., 2014). The speed at which divergence occurs and the resulting genomic landscape depend on the strength of these selective and neutral forces (including drift, mutation and gene flow) occurring within and among populations (Nosil et al., 2009). Across the genome, loci under differential selection in single or multiple populations are expected to diverge first with neutrally evolving genomic regions typically differentiating more slowly (Walsh et al., 2019). Demographic history, such as previous bottlenecks or inbreeding, may also lead to radical changes in the genome, but the effects of these events are expected to occur more evenly throughout the genome (Kimura, 1991; Nei, 2005; Quilodrán et al., 2020). Genomic approaches provide an opportunity to study the genetic landscape of divergence among populations, enabling the relative importance of differing evolutionary forces driving divergence to be determined (Bergström et al., 2022; Estandía et al., 2023; Ravinet, Faria, et al., 2017; Stajich & Hahn, 2005).

Upon colonisation of new environments, populations may be exposed to novel selective pressures, which may result in rapid ecological and phenotypic divergence between populations (Lamichhaney et al., 2015; Walsh et al., 2019). Selection may then act on standing genetic variation or de novo mutations (Estandía et al., 2023). When a locally beneficial allele arises at a locus, positive selection may cause it to rapidly increase in frequency in one population, resulting in a local selective sweep, while in the other population that allele may be lost through drift or purifying selection, or remain at low frequencies (Hejase et al., 2020; Ravinet, Faria, et al., 2017). During the initial stages of divergence genetic differentiation is expected to be localised with peaks of divergence around selected loci, often referred to as 'genomic islands of divergence' (Burri et al., 2015; Ellegren et al., 2012; Nadeau et al., 2012; Nosil et al., 2009). Such areas can be responsible for the accumulation of genetic and phenotypic differences between populations, which may play a fundamental role in speciation (Ruegg et al., 2014; Via & West, 2008). As well as genomic islands of divergence, highly conserved genomic regions – where differentiation is far below background levels – have been identified in a range of species (Hofer et al., 2012; Ravinet, Yoshida, et al., 2017; Sendell-Price et al., 2020; Van Doren et al., 2017). These 'genomic valleys of divergence' may occur because the same allele is favoured in both populations (parallel or balancing selection) (Nielsen, 2005; Roesti et al., 2012; Spurgin & Richardson, 2010) or linked neutral loci are favoured through background selection (Cvijović et al., 2018), leaving a distinct signature of reduced genetic diversity in both populations (Roesti et al., 2014). Furthermore, reduced recombination and background selection can also result in the formation of genomic islands through linkage to selected loci (Noor & Bennett, 2009). Individual-based modelling can be used to further determine if observed patterns align with expectations for non-neutral divergence or if they likely result from stochastic events (Ali et al., 2023; Sendell-Price et al., 2020).

Linkage disequilibrium may facilitate divergence hitchhiking of neutral (and weakly selected or deleterious) loci in proximity to strongly selected loci, resulting in broad genomic islands surrounding these loci (Maynard-Smith & Haigh, 1974; Nosil et al., 2009; Nosil & Feder, 2012; Via, 2012). Broad peaks of divergence often form as a result of recent selection, where recombination has not yet acted upon the regions around the selected loci. Over time recombination erases the effect of divergence hitchhiking, by reducing linkage, which may result in sharp peaks surrounding selected loci (Nosil et al., 2009). However, recombination does not act evenly across the genome and population processes including limited gene flow and small effective population sizes (N_e) are associated with reduced recombination rate, which in turn maintains large regions of divergence (Feder & Nosil, 2010). Discerning and dating the combination of evolutionary processes that have shaped genetic diversity between populations is complex, particularly since the same pattern has the potential to evolve as a result of differing evolutionary scenarios. Therefore, studies need to combine observations of the divergence landscape with knowledge of past and present population processes to determine the relative roles of drift and selection in the divergence of a particular genomic region (Nosil & Feder, 2012; Ravinet, Faria, et al., 2017).

Newly established populations may also be strongly influenced by founder effects, which result in a loss of genetic diversity (Barton & Charlesworth, 1984; Berry, 1986; Harrison, 1991). The characteristics of divergence over time are then mediated by the level and timing of gene flow between populations (Delmore et al., 2020; Räsänen & Hendry, 2008; Ravinet, Faria, et al., 2017), although new species can emerge across differing levels of genetic isolation (Bay & Ruegg, 2017; Kirkpatrick & Ravigné, 2002; Li et al., 2010; Martin et al., 2013). Genetic drift may also act particularly strongly in small, genetically isolated, populations which can result in the rapid loss or fixation of genetic variants (Alleaume-Benharira et al., 2006). The impact of such processes may be exaggerated by sequential founder events and the cumulative effects of drift through a chain of colonisation events (Clegg et al., 2002; Prugnolle et al., 2005; Tomozawa et al., 2014).

The search for genes associated with selection has often involved using genomic patterns of F_{ST} – a measure of the relative difference in allele frequencies between populations. However, since the mechanisms underlying differentiation between populations can be complex, contrasting different measures of genetic diversity can be helpful for inferring differing modes of divergence between populations (Delmore et al., 2018; Irwin et al., 2018; Osmond & Coop, 2020), and a range of diversity statistics can be used (Reviewed in Wolf & Ellegren, 2017), as well as individual-based modelling to disentangle neutral and selective drivers of divergence (Ali et al., 2023).

Island populations of Berthelot's pipit (*Anthus berthelotii*) and its mainland sister species, the tawny pipit (*Anthus campestris*), provide an excellent system to explore genomic patterns of divergence and speciation across divergence timescales and known colonisation events (Armstrong et al., 2018; Illera et al., 2007; Martin et al., 2023;

Spurgin et al., 2014). The ancestor of these two species colonised the Canary Islands from mainland Africa ca. 2.1 million years ago (Mya) and dispersed independently to both the Madeiran and Selvagens archipelagos ca. 50 thousand years ago (kya) and 8 kya respectively (Martin et al., 2023). Previous research suggests colonisation is associated with reduced genetic diversity through founder effects and an absence of post-colonisation gene flow across the Berthelot's pipit range (Gonzalez-Quevedo et al., 2015; Spurgin et al., 2014). This is confirmed by signatures of inbreeding – runs of homozygosity (ROH) > 1 Mb – across the recently colonised archipelagos (Martin et al., 2023). Across the species' range, strong genetic structure exists between, but not within, Berthelot's pipit populations at the archipelago level (Armstrong et al., 2018; Martin et al., 2021).

Importantly, across the Berthelot's pipit system there are strong selection gradients, including climate, habitat and pathogen regimes (Illera et al., 2016). For example, different populations have considerable, and temporally consistent, variation in the prevalence of avian pox and malaria (Illera et al., 2008; Spurgin et al., 2012), which has enabled previous studies of host–pathogen evolution (Gonzalez-Quevedo et al., 2014, 2015; Sheppard et al., 2022). Such pathogens can exert strong selective pressures on avian populations (Liao et al., 2017). There are also significant morphological differences across the system, with reduced body and bill size in Berthelot's pipit compared to the tawny pipit, and archipelago-level variation in bill morphology and body size in Berthelot's pipit (Armstrong et al., 2018; Spurgin et al., 2014). Berthelot's pipit is classified into two subspecies: *Anthus berthelotii berthelotii* inhabits the Canary Islands and Selvagens, while *Anthus berthelotii madeirensis*, which inhabits the Madeiran islands, is characterised by longer bill lengths (Arctander et al., 1996; Illera et al., 2007) and larger body size (Spurgin et al., 2014). The relative role of selection and founder effects in shaping divergence for this trait is currently not clear (Armstrong et al., 2018).

Reduced representation sequencing (RAD-seq) markers have previously been used to investigate selection across the Berthelot's pipit range, both at broad (Armstrong et al., 2018) and fine (Martin et al., 2021) geographical scales. The strongest signatures of selection were identified between archipelagos, compared to between island populations within archipelagos. Loci near to or within genes associated with immunity, metabolism and bill length were identified as being divergent. However, it is not clear to what extent patterns of diversity in these genomic regions are shaped by drift and selection, and it is likely many divergent loci in the genome have gone undetected as a result of unsequenced genomic regions.

To uncover loci of importance for divergence and adaptation, we use whole-genome resequencing to assess genomic landscapes of divergence across sequentially colonised archipelagos by the Berthelot's pipit and its ancestor. Our specific aims were: (1) to determine how divergence has accumulated across the genome between Berthelot's pipits and its mainland relative the tawny pipit (ca. 2.1 Mya), as well as between more recently divergent archipelago populations of Berthelot's pipit (colonised ca. 50 and 8 kya); (2) to identify genomic islands and valleys of divergence and (3) to understand how

drift and selection have interacted to shape variation across these genomic regions. To generate hypotheses about potential adaptive phenotypes, we identify candidate genes in regions under selection.

2 | METHODS

2.1 | Avian sampling, genome resequencing and variant calling

Berthelot's pipit was sampled across the three Macaronesian archipelagos of its range, to which it is endemic (the Canary, Madeiran and Selvagens archipelagos; see Figure 1). These islands have formed as a result of volcanic activity 1–26 million years ago (Florencio et al., 2021) and support a range of habitats across which the Berthelot's pipit is relatively abundant (Garcia-del-rey & Cresswell, 2007). We used Berthelot's pipit blood samples ($n=11$) from individuals across six island populations (spanning the three archipelagos) to maximise the geographical range of samples and to account for differences in habitat, pathogen exposure and demographic history variation across the species (Table 1). One male and female sample was selected for each population, and all individuals had no detected malaria infection when screened for *Haemoproteus* and *Plasmodium* spp. or evidence of pox lesions (Illera et al., 2008). In addition, we sequenced one sample from the Berthelot's pipit mainland sister species, the tawny pipit (*Anthus campestris*, sampled Latitude: 17.991703°, Longitude: – 146 16.016672°). This allowed a range of divergence levels and colonisation timeframes to be compared within Berthelot's pipit and its sister species (Table 1). Blood samples were stored in absolute ethanol (800 μ L) at 4°C and DNA was extracted using a salt extraction protocol (Richardson et al., 2001), and individuals were molecularly sexed (Griffiths et al., 1998).

We used a draft Berthelot's pipit reference genome assembly, generated by Armstrong et al. (2018), to align genome-wide sequence reads and to call genomic variants (see below). This assembly was generated using a male bird from Porto Santo in the Madeiran archipelago and was selected based on having low genome-wide heterozygosity as determined by previous RAD-seq analyses (Armstrong et al., 2018). This genome was assembled de novo from paired-end Illumina reads (2 × 125 bp on HiSeq 2500 sequencer) using DISCOVAR (Weisenfeld et al., 2014). The resulting draft assembly has a total size of 1.15 Gb (94.3% of the length of the Zebra finch genome), with a BUSCO score (Simão et al., 2015) of 64% (vertebrate-specific single-copy orthologues) and contig N50 of 355 kb. For full details see Armstrong et al. (2018).

To perform whole-genome resequencing and variant calling against the Berthelot's pipit genome assembly, low input transposase-enabled libraries were sequenced across four lanes using paired-end Illumina sequencing on a HiSeq 4000 (2 × 150 bp), generating high genome read coverage (>20× per individual; Table S1). Raw reads were merged at the individual level, trimmed (Phred quality score > Q30) and aligned to the draft Berthelot's pipit genome assembly using BWA-mem v0.7.12 (Li, 2013). Potential duplicate reads

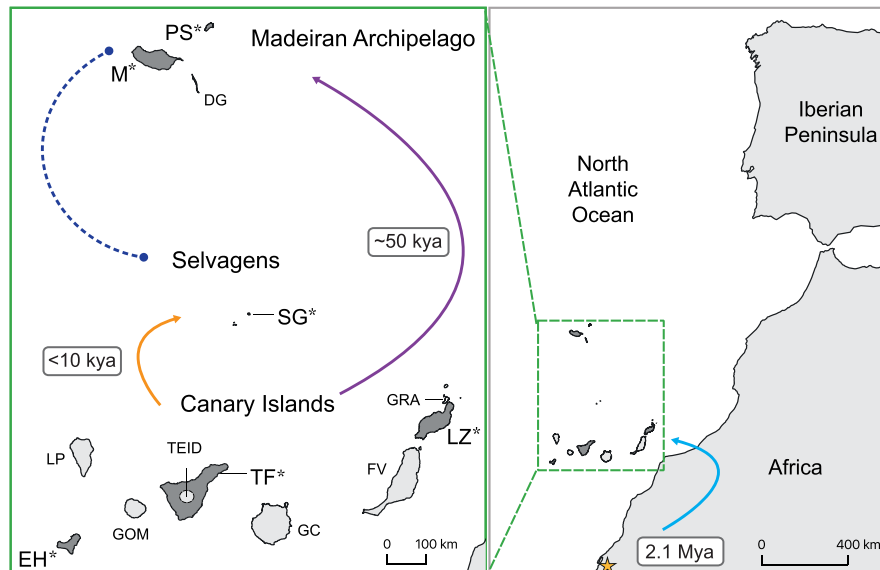


FIGURE 1 Map of the island endemic Berthelot's pipit populations across its Macaronesian range and its ancestor, the tawny pipit, sampling location in Mauritania (see star). Populations included in this study are denoted with an asterisk and the island shaded dark grey. Solid lines with arrows indicate population comparisons, with the timing and direction of colonisation events highlighted (Martin et al., 2023). The blue dotted line indicates comparison between Selvagens and Madeiran populations after these were colonised in two independent events from the Canary Islands. Canary Island populations: El Hierro (EH), La Palma (LP), La Gomera (GOM), El Teide (TEID) mountain population on Tenerife (>2000 m a.s.l.), Tenerife (TF), Gran Canaria (GC), Fuerteventura (FV), Lanzarote (LZ), La Graciosa (GRA). Madeiran populations: Madeira (M), Porto Santo (PS) and Deserta Grande (DG). Selvagens: Selvagem Grande (SG).

TABLE 1 Sampling information for divergence comparisons across populations of Berthelot's and tawny pipit.

Divergence comparison	Population locations	Estimated divergence timeframe	Bottleneck severity	n retained loci
Tawny pipit, sister species/Canary Islands	Mauritania, Mainland Africa/ LZ, TF & EH	2.1 Mya ^a	Weak. Genome-wide reduction in diversity	10,829,660
Canary Islands/Madeiran archipelago	LZ, TF & EH/M & PS	50 kya ^a	Strong. Founder effect, contemporary population recovery	5,590,607
Canary Islands/Selvagens archipelago	LZ, TF & EH/SG	8–10 kya ^a	Strong. Founder effect, contemporary inbreeding	5,266,205
Madeiran archipelago/Selvagens archipelago	M & PS/SG	50 kya ^a	Very strong. Two independent bottlenecks	3,733,990

Note: Population codes from: (1) Canary Islands: Lanzarote (LZ), Tenerife lowland (TF), El Hierro (EH); (2) Madeiran archipelago: Madeira (M), Porto Santo (PS) and (3) Selvagens archipelago: Selvagem Grande (SG). n retained loci = autosomal and Z mapped loci retained for each population comparison.

^aDivergence timeframes from Martin et al. (2023).

were then marked using Picard tools *MarkDuplicates* in GATK v4.1 (McKenna et al., 2010), and binary alignment files (.bam) were validated prior to variant calling. GATK *HaplotypeCaller* was then used in GVCF mode to call genomic variants, with the Berthelot's pipit draft genome assembly using as a reference to call variants against. Unmapped or poor-quality reads (with a root-mean-squared read mapping quality <25) were discarded. To account for errors in read mapping, variants were then filtered for read strand bias (Fisher's exact test >60 and Strand Odds Ratio >3) and quality by depth (QD <2) using GATK. Joint genotyping was then performed across samples, to reduce call error rate, using GATK's *GenomicsDBImport*

and *GenotypeGVCF* tools. Contig-level Variant Call Format (VCF) files were combined using GATK *SortVcf*, with variants mapped to contigs less than 500 bp removed.

To allow chromosome-level genomic divergence to be assessed and visualised, genomic variants were mapped to relative chromosome positions of the Zebra finch (*Taeniopygia guttata*) genome assembly bTaeGut1_v1.p (NCBI Assembly GCA_003957565.1) (Warren et al., 2010) using the SatsumaSynteny module within Satsuma (Grabherr et al., 2010), which performs well on fragmented genome assemblies (Liu et al., 2018). Since our draft Berthelot's pipit reference genome is a highly fragmented

assembly, this collinearity mapping allowed pseudochromosome locations to be determined which could be used for assessing genomic landscapes of divergence and to make use of gene identities in the well annotated Zebra finch genome. To visually assess the collinearity of the Berthelot's pipit and Zebra finch genome assemblies, we used the D-GENIES dot-plot tool (Cabanettes & Klopp, 2018) with the default options. This showed that high synteny exists between these genomes, although percentage identity is relatively low for most contigs (25%–75%). Despite this, only a very few short regions of the Berthelot's pipit genome are misassembled/misplaced, many of which are on the only assembled sex chromosome (Figure S1). Satsuma determines the order, location and orientation of Berthelot's pipit contigs relative to Zebra finch chromosomes. This output was used to reassign contig locations of genomic variants to pseudochromosome locations using custom R scripts (RStudio Team, 2016).

The VCF files were generated to assess (1) divergence through speciation between the tawny and Berthelot's pipit and (2) divergence among Berthelot's pipit populations across archipelagos. For the first dataset 'All Pipits' variants were joint called across the Berthelot's and tawny pipit individuals. For the second dataset 'Berthelot's' variants were joint called across only the Berthelot's pipit individuals. Both datasets were filtered in VCFtools v0.1.15 (Danecek et al., 2011); removing unmapped sites (--not-chr 0), indels (--remove-indels), sites with >2 alleles (--max-alleles 2), sites with more than four failed genotype calls (--max-missing-count 4) and variants with accuracy <99.9% (--minQ 30). Sites were further filtered to remove those at which the mean read depth < 10 or more than twice genome-wide average (Li, 2014) (>45 for 'All Pipits', >44 for 'Berthelot's') (--min-meanDP 10, max-meanDP 45/44). Quality and coverage statistics for both datasets are provided in Table S1.

2.2 | Population comparisons

We compared pairwise divergence between the tawny and Berthelot's pipit, and then among three dyads of Berthelot's pipit archipelago populations (Table 1, Figure 1). The dyads chosen for comparison varied in their temporal, spatial and morphological divergence, and in terms of selective pressures as outlined below:

2.2.1 | Tawny pipit versus Canary Islands

The ancestor of Berthelot's pipit initially colonised the Canary Islands ca. 2.1 Mya (Martin et al., 2023; Voelker, 1999) with founder effects resulting in a genome-wide reduction in genetic diversity compared to the tawny pipit (its sister species). There is no evidence of subsequent gene flow. The tawny pipit is a Palearctic migrant. After ending the breeding season Western European populations winter mainly in sub-Saharan Africa, while breeding Asian populations winter in the Arabian Peninsula and Southwest Asia (Tyler &

Christie, 2020). In contrast, Berthelot's pipit is an island resident. Habitat types and disease prevalence varies substantially across the Berthelot's pipit Canary Islands range. Both species are exposed to avian malaria, avian pox and other pathogens (tawny pipit; Calero-Riestra & Garcia, 2016, Berthelot's pipit; Illera et al., 2008).

2.2.2 | Canary Islands versus Madeiran archipelago

Colonisation of the Madeiran archipelago from the Canary Islands by Berthelot's pipit is estimated to have occurred ca. 50 kya (Table 1, Figure 1; Martin et al., 2023), and resulted in a strong population bottleneck. Berthelot's pipit from the Madeiran archipelago are classified as a separate subspecies *A. berthelotii madeirensis* (Hartert, 1905). The Madeiran populations show the longest bill lengths (Illera et al., 2007) and larger body size (Spurgin et al., 2014). Considerable variation in disease prevalence occurs within and among these two islands archipelagos (Illera et al., 2008).

2.2.3 | Canary Islands versus Selvagens archipelago

The Selvagens was populated by Berthelot's pipit through colonisation from the Canary Islands ca. 8–10 kya (Table 1, Figure 1; Martin et al., 2023). Small island size (ca. 3 km²), geographic isolation and strong founder effects have resulted in low genetic diversity in the Selvagens' population and strong signatures of inbreeding (Martin et al., 2023; Spurgin et al., 2014). While there is considerable variation in disease prevalence across the Canary Islands, no disease has been detected on the Selvagens (Illera et al., 2008).

2.2.4 | Madeiran archipelago versus Selvagens archipelago

Berthelot's pipit populations across these two archipelagos are separated by ca. 50 thousand years and two independent bottleneck events, with no evidence of post-colonisation gene flow between the archipelagos (Illera et al., 2007; Martin et al., 2021; Spurgin et al., 2014). While Berthelot's pipit within the Selvagens are not infected with avian pox or malaria, there are strong differences in disease prevalence between islands within the Madeiran archipelago (Illera et al., 2008).

2.3 | Differentiation landscapes and genomic islands

We calculated pairwise F_{ST} between the tawny pipit and the Canary Island Berthelot's pipit population (one comparison) using the 'All Pipits' dataset, and between the three Berthelot's pipit archipelago populations using the 'Berthelot's' dataset in VCFtools (Table S1). Variation across the genome was visualised using Manhattan plots.

To identify highly divergent genomic regions between population comparisons, pairwise F_{ST} values were Z-transformed. Since population history shapes genetic variation between populations, baseline levels of divergence vary significantly between each of the population comparisons (see Section 3). Within each comparison, we classified windows as divergent if their mean F_{ST} was more than five standard deviations greater than the genome-wide mean ($zF_{ST} > 5$) and in the top 1% of SNP windows, which is a conservative approach to identifying outliers (Choi et al., 2020; Han et al., 1999; Lamichhaney et al., 2015; Walsh et al., 2019). For each population comparison, F_{ST} was calculated in 50kb non-overlapping windows. Genomic linkage typically extends 25–35kb in the three archipelago populations of Berthelot's pipit (Martin et al., 2021). We used a 50-kb window for genomic island detection because it provided sufficiently fine resolution across the genome while representing 184–532 sites per window. Finally, windows containing <30 sites were removed prior to conducting analysis (~1% of windows). We considered the Z chromosome separately due to known differences in evolutionary pressures across sex chromosomes (Ellegren, 2011), and poor pseudochromosome assembly quality (Figure S1). It is also important to note that the exact number of genomic islands of divergence per chromosome may differ if there was a chromosome-level pipit assembly.

By investigating genomic variation across the tawny pipit and Berthelot's pipit speciation event, it was also possible to identify highly conserved (low divergence) genomic regions which may have a role in parallel adaptation between the species (Van Doren et al., 2017). To do this we applied the same zF_{ST} approach, instead identifying windows with mean F_{ST} less than 5 standard deviations below the genome-wide mean.

2.4 | Detailed characterisation of variation in divergence peaks

To identify regions putatively under selection in elevated regions of differentiation, we compared values of Tajima's D and nucleotide diversity (π) in and outside of outlier windows (i.e. assessing whether regions of elevated differentiation had corresponding dips in Tajima's D and π). We calculate 50kb-windowed Tajima's D and π using VCFtools, for each population dyad to allow direct comparison. Tajima's D can be used to test for deviations from neutral evolution (values near or equal to zero), with positive values indicative of balancing selection or sudden population contraction, and negative values indicative of a recent selective sweep or population expansion following a recent bottleneck (Tajima, 1989). Nucleotide diversity (π) is defined as the number of nucleotide differences per site between sequences within a population. Estimating π across the genome may reveal population-level diversity within genomic regions, which can be used together with Tajima's D to make inferences about potential evolutionary forces acting within regions of interest.

To identify genes located within divergent windows, we viewed 50kb regions of interest using the Zebra finch genome (v. bTae-Gut1_v1.p) in NCBI Genome Data Viewer v. 4.8. (www.ncbi.nlm.nih.gov/genome/gdv/browser). Patterns of divergence across peak regions were assessed across the three archipelago dyadic comparisons. Where several windows exceeded this threshold within a genomic island, we assessed the distribution of F_{ST} within peaks, and where appropriate highlight the most likely candidate genes under selection.

2.5 | Simulations of divergence

We performed individual-based simulations of genomic divergence, under neutral and selective processes using the 'glads' R package (Quilodr  n et al., 2020). These simulations consequently provide a comparison of neutral and non-neutral divergence patterns, which we used to compare to our empirical genome-wide F_{ST} distributions.

The glads model, with an additive genotype–phenotype map, simulated the divergence of two populations – parameterised to represent the demography and colonisation history of Berthelot's pipit in the Canary Islands and Selvagens archipelagos. Each population comprised of 300 biallelic SNPs, with 50 loci under selection and the remainder neutral (i.e. did not contribute to the additive phenotype, but were influenced by neutral evolutionary processes such as drift and recombination). Simulations excluded post-divergence gene flow (Martin et al., 2021), with population sizes set at 4000 and 400 diploid individuals for the Canary Islands and Selvagens respectively (Spurgin et al., 2014). A recombination rate of 3cM/MB was used, based on the genome-wide average of the collared flycatcher (*Ficedula albicollis*) for a 75-Mb chromosome. Mutation and migration were not included in the simulation. Assuming a divergence time of 8000 years and a generation time of 2.2 years (Martin et al., 2023), 3636 generations were simulated. The simulation was repeated independently 100 times for both parallel and divergent selection, maintaining the same starting parameters, and ensuring distinct genetic identities for individuals (i.e. genotype for the 300 loci) at the onset of each independent simulation.

3 | RESULTS

3.1 | Whole-genome resequencing

Illumina sequencing resulted in 1,030,115,042 paired-end reads, providing mean individual read coverage of 23.6 ± 2.6 s.d. when mapped to the Zebra finch's 1.1 Gb genome (Warren et al., 2010). Variants were called, mapped to the Zebra finch assembly and quality filtered, resulting in 11,788,225 mapped SNPs in the 'All Pipits' dataset and 5,934,934 in the 'Berthelot's' dataset. Individuals had low levels of missing data (<5%). Final depth of coverage for the quality filtered SNPs was high (Jiang, Jiang, et al., 2019; Sims et al., 2014),

with a mean 24.3 X for the 'Berthelot's' dataset and 24.7 X for the 'All Pipits' dataset (Table S1).

3.2 | Distributions of F_{ST} between comparisons

Autosomal mean F_{ST} was high between the tawny and Berthelot's pipit (Figure 2; Martin et al., 2023) but low between Berthelot's pipit archipelago populations separated by one colonisation event (i.e. Canary Islands vs. Madeira, Canary Islands vs. Selvagens). The distribution of F_{ST} was positively skewed between the Canary Islands and both subsequently colonised archipelagos (Madeira and Selvagens) but approaches a normal distribution between the tawny and Canary Islands Berthelot's pipit. Between single colonisation events, standard deviations of windowed F_{ST} values were low, ranging 0.041–0.070. The spread of divergence scores was much greater between the Selvagens and Madeiran archipelago (s.d.=0.139), which are separated by two independent colonisation events (Figure 2).

Like the autosomes, the Z chromosome showed increasing divergence over longer timeframes between single colonisation events, with highest divergence between the tawny pipit and Berthelot's pipit and lowest between the most recently separated Berthelot's pipit archipelago populations (i.e. the Canary Islands and Selvagens; Figure S2a). Between the tawny and Berthelot's pipit, mean F_{ST} divergence across the Z chromosome exceed that observed across the autosomes, while the converse was found across Berthelot's pipit archipelago populations (Table 2).

3.3 | Overall correlations between genomic landscapes of divergence

Divergence at genomic loci in tawny-Berthelot's pipit comparisons was significantly but weakly correlated with divergence at these loci between Berthelot's archipelago comparisons (Figure 3 and Figure S3, Pearson's correlation: $r = .02-.15$, $p < .01$). As expected,

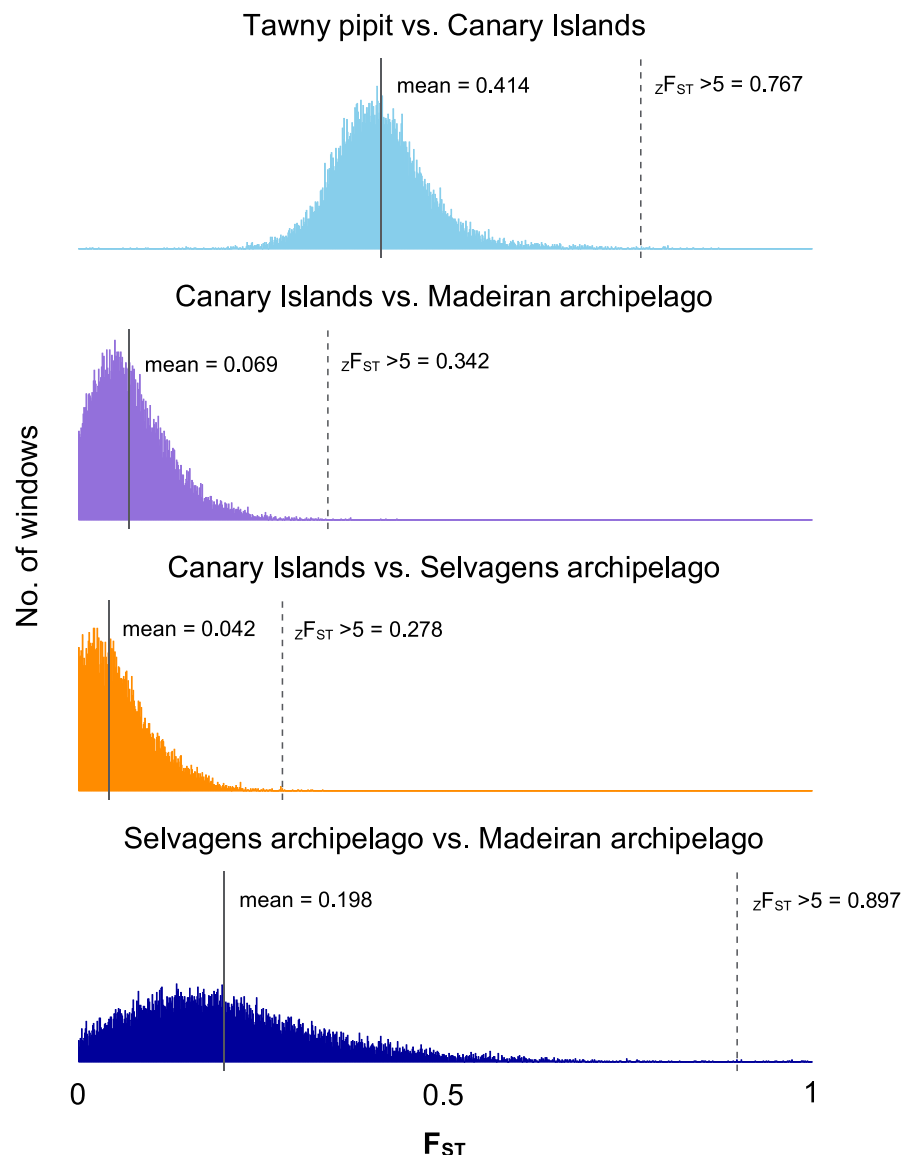


FIGURE 2 Distribution of pairwise genomic differentiation, F_{ST} , across Berthelot's and tawny pipit comparisons calculated in 50 kb autosomal windows. Positions of means (solid line) and $zF_{ST} > 5$ threshold (dotted black line) are highlighted.

TABLE 2 Genomic differentiation, F_{ST} , between Berthelot's pipit and tawny pipit, and among Berthelot's pipit archipelagos.

Comparison	Autosomal mean F_{ST}	Z Chr mean F_{ST}	Autosomal $zF_{ST} > 5$	Top 1% SNPs	n windows $zF_{ST} > 5$
Tawny pipit, sister species/Canary Islands	0.414	0.427	0.767	0.637	27
Canary Islands/Madeiran archipelago	0.069	0.063	0.342	0.229	9
Canary Islands/Selvagens archipelago	0.042	0.034	0.278	0.186	10
Selvagens archipelago/Madeiran archipelago	0.198	0.171	0.897	0.609	22

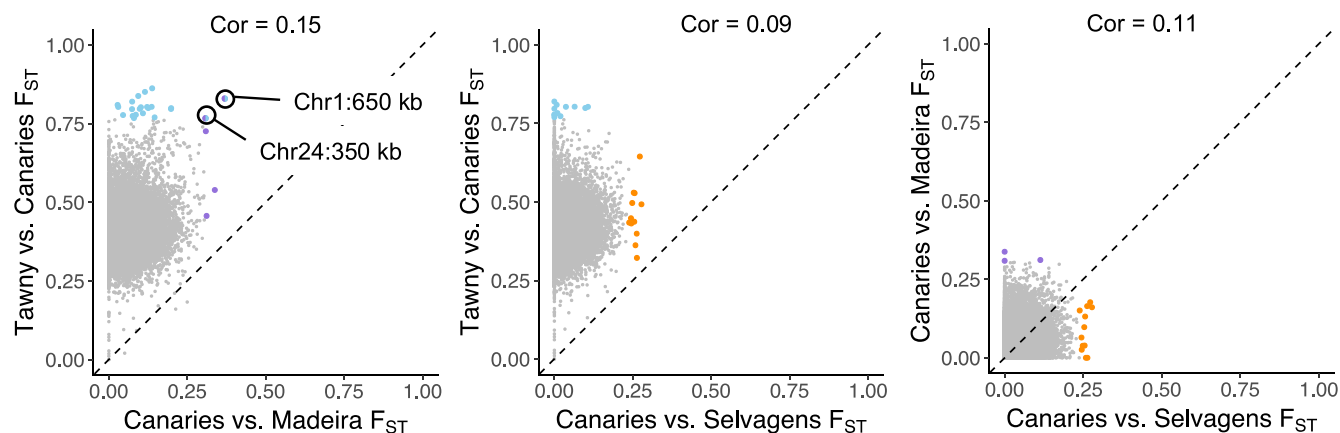


FIGURE 3 Correlated patterns of F_{ST} between Berthelot's pipit population comparisons, and between Berthelot's and tawny pipit populations separated by single colonisation events. F_{ST} was calculated in 50kb non-overlapping windows, outliers ($zF_{ST} > 5$) are highlighted; Tawny versus Canary Islands=light blue, Canary Islands versus Madeira=purple, Canary Islands versus Selvagens=orange. The only two outlier windows which occur in multiple population comparisons are circled and the chromosomal positions identified.

the divergence of loci between dyadic comparisons was more strongly correlated when those comparisons involved a common population (because of the characteristics of loci in that common population) and not because of independently reoccurring divergence of the same loci. The highest correlation explains 34.8% of variation in dyadic F_{ST} scores (Pearson's correlation: $r = .59$, $p < 2.2 \times 10^{-6}$, Canary Islands vs. Madeira compared to Selvagens vs. Madeira).

3.4 | Identification of peaks and valleys and correlations between population comparisons

The genomic landscape of divergence between the tawny pipit and the Canary Islands Berthelot's pipit population was broadly homogeneous across the genome, with 27 windows > 5 s.d. from the genomic autosomal mean ($F_{ST} > 0.41$) (Figure 4, Table 2). The distribution of

these areas of elevated divergence was non-random with 12 well defined clusters of high F_{ST} , seven of which exceeded the criteria to be considered 'divergence peaks' or 'genomic islands of divergence' (Figure 4, Table S2). Peak size was in the range of 0.05–2 Mb (Table S2) and such divergence peaks covered 0.14% of the genome. Across the Z chromosome, two broad peaks (1–2 Mb) of strong divergence are also observed between the tawny and Canary Island Berthelot's pipit, although we stress that these patterns may be as a consequence of poor contig mapping on this chromosome. The pattern of divergence within the different genomic islands fell into two categories: (i) most commonly, regions of sharp divergence usually include one or two peaks where a "top peak" with the highest associated F_{ST} could be identified or (ii) a broad peak of similar F_{ST} divergence across the island (see Figure 5). Only one such broad peak of divergence was identified approximately 2 Mb in length on chromosome 1A (Figure 5a), with the strongest F_{ST} divergence scores across the genome.

FIGURE 4 Pairwise genomic differentiation, F_{ST} , across the genome for Berthelot's and tawny pipit comparisons calculated in non-overlapping 50kb windows. Genomic islands of divergence or valleys of similarity, where F_{ST} is 5 standard deviations greater than, or less than, the mean window value, are highlighted (indicated by the dashed horizontal line). Pseudochromosomes (derived by comparison with chromosomes of the Zebra finch genome) are shown in alternating light and dark shading. Vertical coloured bars indicate the location of genomic islands within each population comparison. Labelled arrows indicate the location of shared genomic islands and the genomic distance between closely located peak windows across population comparisons. Population comparison highlights: light blue = Tawny versus Canary Islands, Canary Islands versus Madeiran islands = purple, Canary Islands versus Selvagens = orange and Selvagens versus Madeira = navy.



Autosomal divergence between Berthelot's pipit archipelagos showed different patterns. Archipelago comparisons with just one founding step separating them (Canary Islands vs. Madeira and

Canary Islands vs. Selvagens) were characterised by low genome-wide divergence with few strongly differentiated 'islands of divergence' (Figures 2 and 4). In contrast, the divergence landscape was

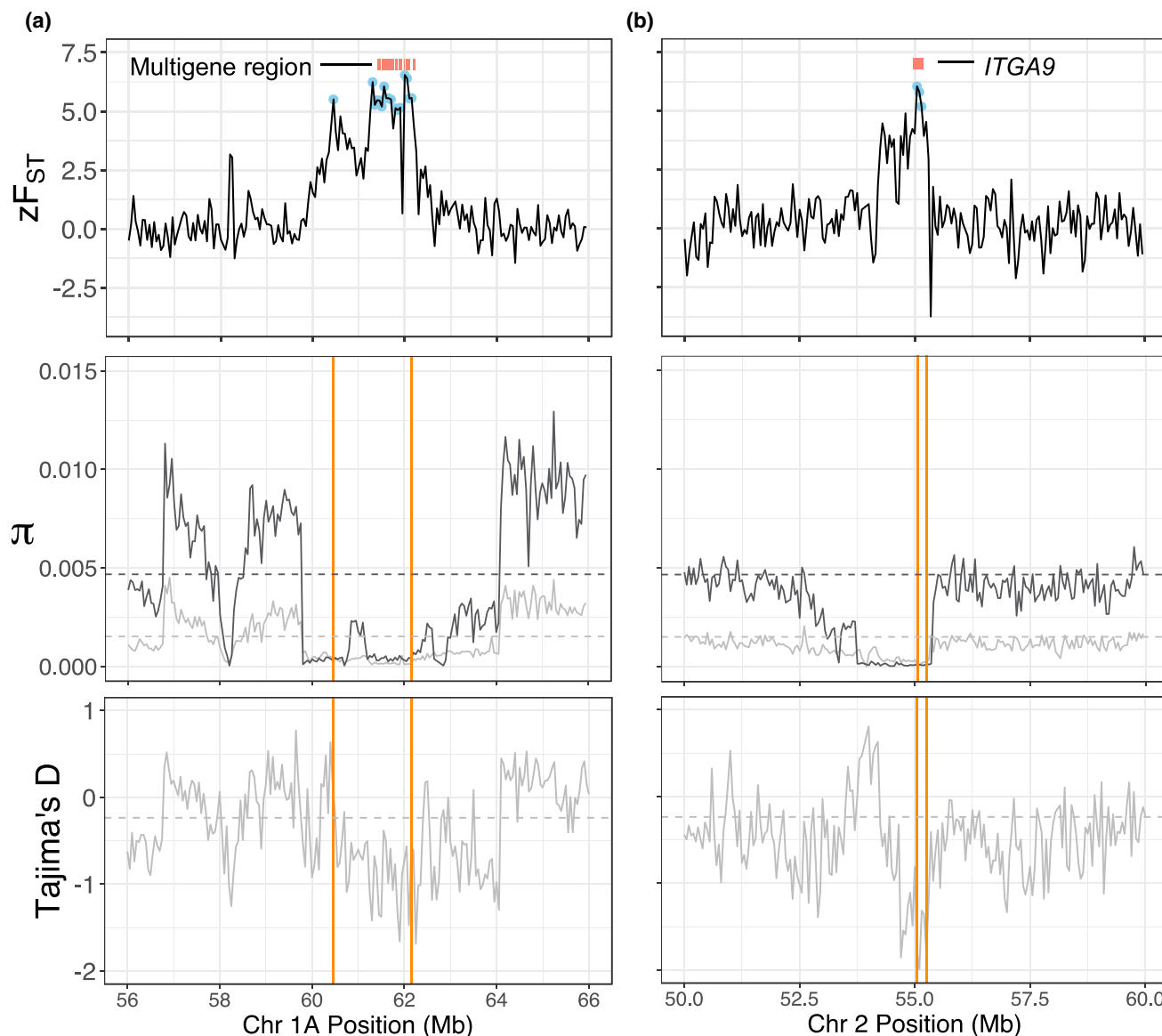


FIGURE 5 Patterns of divergence (zF_{ST}), genetic diversity (π) and a measure of the loss of rare alleles (Tajima's D) surrounding two genomic islands of divergence between the tawny pipit and Canary Islands Berthelot's pipit comparison. These regions are presented as examples to demonstrate (a) a broad region, or 'plateau' of elevated divergence and (b) a narrow peak of divergence. Values were calculated in 50kb windows, with $zF_{ST} > 5$ highlighted in the first panel in light blue, with corresponding genomic locations of the peak start and end indicated by vertical orange lines in panel 2 and 3. Candidate gene locations within peaks are indicated; details in Table S2. Tajima's D and π are reported for the Canary Islands (light grey) and the corresponding π indicated for the tawny pipit (dark grey), with autosomal averages indicated by horizontal dotted lines.

highly heterogeneous when the two independently bottlenecked archipelagos, Selvagens versus Madeira were compared. We identified 9–22 genomic islands of divergence in each Berthelot's pipit archipelago comparison (Figure 4, Table 2), which represented 0.05%–0.1% of 50kb windows. In contrast to the tawny versus Berthelot's pipit comparison, Z chromosome divergence was consistently (but marginally) lower than autosomal divergence between Berthelot's pipit archipelagos comparisons (Table 2), and no strongly divergent genomic windows were identified.

The position of divergent windows varied between population comparisons, with only two shared windows identified between

multiple population comparisons (Figure 3). These two windows, mapped to regions of chromosome 1 and 24, and were shared between the tawny pipit versus Canary Island Berthelot's pipit population comparison and between Canary Island Berthelot's pipit versus Madeiran archipelago comparison. No windows of strong divergence were shared between the different Berthelot's pipit comparisons. Between the tawny and Berthelot's pipit, the majority of the genome exhibited strong divergence with the highest window having $F_{ST} = 0.87$, and only two windows where F_{ST} was zero (Figures 2 and 4). However, we identified seven genomic regions with low divergence, more than five standard deviations

below the genome-wide mean, which mapped to five distinct chromosome regions (Figure 4). Between Berthelot's pipit archipelagos comparisons many 50kb regions of the genome were not divergent ($F_{ST}=0$, Figure 4).

3.5 | Genes in peaks and valleys and patterns of diversity

The identified regions of elevated divergence between the tawny and Berthelot's pipit consistently exhibit corresponding dips in Tajima's D and π in the Canary Islands Berthelot's pipit population, indicative of selective sweeps. Patterns of reduced Tajima's D and π , and elevated F_{ST} were consistent across the broad peak on chromosome 1A, suggesting strong linkage disequilibrium in this region, while the other strongly divergent peaks were narrow in width. Across all strongly divergent regions, six of seven peaks harboured protein-coding sequence of which five regions had named candidate genes (Table S2).

A high number of divergence windows were located within the extended genomic island mapped to chromosome 1A (17 of 27 windows identified between the tawny and Berthelot's pipits). This region contained 17 annotated genes, of which at least seven are associated with immune response, three with craniofacial development and two with metabolism (see Table S2). The single most strongly divergent peak in the region included two genes: *CMA5* associated with the innate immune response (O'day et al., 2018; Urbanek et al., 2020) and *ABCC9*, associated with cartilage and bone development (Czeschik et al., 2013) (Table S2).

To further investigate divergence across this broad region on chromosome 1A, we investigated patterns of divergence between the tawny pipit and all Berthelot's pipit populations, and among Berthelot's pipit populations. The results confirm (i) divergence of this region occurs between the tawny pipit and all Berthelot's pipit archipelagos and (ii) this region exhibits low divergence between Berthelot's pipit archipelagos (Figure S4). The other narrow regions (<150kb) of strong divergence between the tawny pipit and Berthelot's pipit mapped to genomic regions including genes putatively associated with immune response and wound healing (Cui et al., 2017; Sigurðarson, 2020), development of the retina (Xu et al., 2020) and carbohydrate metabolism (Han et al., 1999) (Table S2).

Across the Z chromosome, two broad peaks (1–2Mb) of high divergence occur between the tawny pipit and Canary Islands Berthelot's pipits. These have corresponding dips in nucleotide diversity in both populations (Figure S2b), which suggests they may be under divergent selection. Identified genes within the top 50kb divergent window in both peaks, both have strong association with hearing: *ADGRV1* is associated with hearing loss and retina development (Yan et al., 2018); and *PIP5K2* regulates hearing through growth and maintenance of sensory cells in the inner ear (Yousaf et al., 2018). Nucleotide diversity across the Z chromosome shows a consistent peak across all populations of high diversity at ~45 Mb,

likely an artefact of incorrect mapping of W chromosome reads to the Z in female individuals.

Strongly divergent regions between archipelago populations of Berthelot's pipit had varied patterns of π and Tajima's D across different populations, indicating a range of evolutionary processes occurring within these regions (Table S3). Across the Canary Islands versus Selvagens comparison, several highly divergent regions mapped to genomic locations within previously identified (Martin et al., 2023) long ROH > 1 Mb in the Selvagens individuals (and negative Tajima's D). Across the strongly divergent regions for the Canary Islands versus Madeiran archipelago comparison, Tajima's D was consistently elevated within the Madeiran archipelago.

Genes in islands of divergence across the Berthelot's pipit range are associated with similar biological functions (Tables S2 and S3). Regions under putative selection within the Selvagens include genes associated with craniofacial shape, apoptosis and inflammation, development of the retina and teeth, metabolism, muscle and growth (Table S3). Within the Madeiran archipelago population regions potentially under selection include genes associated with facial, skin and bone development, immunity (innate immune response and regulation of B and T cells) and eye development.

3.6 | Genes with low divergence between species

Between the tawny and Berthelot's pipit, we identify seven low-divergence genomic regions including a total of 11 candidate genes, of which seven were annotated. These likely strongly conserved regions included genes putatively associated with pathogen infection, inflammation and platelet regeneration, growth factor pathways, and muscle and limb development (Table S4). For example, *NFX1* is a transcriptional repressor of major histocompatibility complex (MHC) class II genes (Gewin et al., 2004; Strominger et al., 1994) with biological function in immune and inflammatory response. We also identify a genomic valley with corresponding peaks of nucleotide diversity in Berthelot's pipit (and negative Tajima's D) and low diversity in the tawny pipit, putatively associated with recent population growth across the Canary Islands. This loci was within the genomic region containing *RANBP3* which is associated with growth factor pathways including bone morphogenic protein (BMP) signalling (Dai et al., 2009).

3.7 | Simulations of divergence

Individual-based modelling captured elevated F_{ST} within genomic regions under both divergent and parallel selection relative to neutrally evolving genomic regions (Figure S5). Over the 100 independent simulations, F_{ST} ranged from 0 to 1. The average level of F_{ST} across the 100 independent simulations indicated a propensity for islands to form at selected loci (as shown by the higher average level of F_{ST} at selected loci in contrast to neutral loci; 125–175bp locus position) (Figure S5). Genomic divergence was observed most

strongly under divergent selection compared to parallel selection, with loci displaying less variation in F_{ST} scores consistently across the 100 independent simulations.

4 | DISCUSSION

We find that divergence accumulates across the genome in association with increasing timeframes since population founding in an island colonising bird, confounded by the cumulative effects of drift though sequential founder events. In this study, we have shown that through population founding across archipelagos, genomic divergence accumulates in few loci and linked regions, but over time drift causes genome-wide loci to also become differentiated across the speciation continuum (from recently divergent Berthelot's pipit populations separated 8–50 kya to the reproductively isolated tawny pipit and Berthelot's pipit separated ca. 2.1 Mya). We find only weakly correlated patterns of genetic divergence and different strongly divergent loci between populations through different founder events (i.e. the same loci rarely show strong divergence across different events). However, we do find that the loci within these genomic islands of divergence are consistently (putatively) associated with the same ecologically important traits, that is, body/head size and immune defence, across the range (see [Tables S2 and S3](#)).

4.1 | Divergence across the speciation continuum

By comparing F_{ST} frequency distributions between sequentially founded, geographically isolated populations of Berthelot's pipit, we show how genomic differentiation accumulates across timeframes in the absence of gene flow ([Figure 2](#)). Berthelot's pipit populations in the early stages of divergence (Canary Islands vs. Selvagens) after a recent colonisation event (<10 kya), have a positively skewed F_{ST} frequency distribution. This includes a handful of strongly divergent genomic regions, which may differ due to differential selection or bottleneck effects on genetic drift. Where populations have been geographically separated for longer timescales (i.e. Canary Islands vs. Madeiran archipelago ca. 50 kya) leading to formation of subspecies, drift has had time to cause overall genomic divergence (Nosil et al., 2009). Over much longer divergence timeframes (i.e. across the speciation event between the tawny and Berthelot's pipit separated >2 Mya), divergence accumulates across the whole genome (mean F_{ST} =0.41), approximating a normal distribution with only a few weakly or strongly divergent regions. These patterns are in concordance with speciation models from early- to late-stage divergence with geographical isolation (Seehausen et al., 2014).

Across the Berthelot's pipit system, we can also compare genomic divergence between two independently bottlenecked populations, founded from the same source population (Canary Islands) at different time points (ca. 50 and 8 kya) ([Figure 1](#)). Here, the F_{ST} frequency distribution has a greater spread, likely due to the combined

consequences of two founder events and subsequent drift. Unlike the comparisons of populations separated by one colonisation event, in this scenario regions that have diverged due to natural selection are more difficult to identify due to high levels of overall genome-wide divergence. Our patterns of genomic divergence among the different Berthelot's pipit archipelago populations concur with comparisons of rapidly speciating island silvereye (*Zosterops lateralis*) populations separated by <200 years where Sendell-Price et al. (2021) only identified a few strongly divergent genomic regions. Between-species comparisons of *Ficedula* flycatchers (Burri et al., 2015) and *Saxicola* stonechats (Van Doren et al., 2017), with divergence times of several millions of years, showed similar normal distribution patterns to that which we observed between the tawny and Berthelot's pipit.

4.2 | Independent patterns of divergence through population founding

We identified weakly correlated patterns of F_{ST} across genomes from the different Berthelot's pipit populations, and between Berthelot's pipit and its sister species, the tawny pipit ([Figure 3](#) and [Figure S2](#)). That the same loci/regions are not divergent across population comparisons, suggests that long-term linked selection is not a major reason for divergence in the Berthelot's pipit. Instead, such divergent loci may be due to independent evolutionary responses to selection pressures (Munch et al., 2016). Our findings contrast with several studies of genomic landscapes across divergence timescales and through varied gene flow contexts (Burri et al., 2015; Renaut et al., 2013; Stankowski et al., 2019; Van Doren et al., 2017; Vijay et al., 2017), which show parallel patterns of divergence between geographically and morphologically distinct taxa. Instead, we find only weakly correlated F_{ST} scores and different putatively selected genomic regions diverging (see below) through time and speciation events across the range ([Figures 3 and 4](#)). Why this is the case in Berthelot's pipit remains unknown, but it may be attributed to the stochastic nature of the founding process, as has been reported in, among others, island colonising silvereyes (Sendell-Price et al., 2021) and laboratory range expansion experiments of red flour beetle (*Tribolium castaneum*; Weiss-Lehman et al., 2019). Largely independent patterns of divergence between population pairs are also reported in a range of other systems, for example through parallel speciation of stick insects (*Timema cristinae*; Soria-Carrasco et al., 2014), sympatric environmental adaptation of flatfish species (Le Moan et al., 2019) and adaptive radiation in Lake Victoria *Pundamilia* cichlid fishes (Meier et al., 2018).

4.3 | Genomic islands of divergence and repeated selection for traits

Absence of secondary contact between the tawny and Berthelot's pipit for ca. 2.1 million years has resulted in considerable

genome-wide divergence. Across this speciation event, we identified seven weakly divergent (or conserved) and seven strongly divergent autosomal genomic regions (Figures 2 and 4). The most divergent region formed a broad ~2Mb peak on chromosome 1A. We also identified two peaks of broad divergence across the Z chromosome (Figure S2), although this may be an artefact of poor-quality variant mapping on this chromosome (Figure S1). All other strongly divergent regions formed narrow sharp peaks. Through subsequent independent founder events from the Canary Islands to the Selvagens and Madeira, we identify 12 and five strongly divergent 50kb windows, respectively, all mapped to autosomal chromosomes (Table S3).

Individual-based model simulations showed localised genomic islands of divergence forming around selected loci (Figure S5). Our models parameterised for the most recently divergent Berthelot's pipit populations, show that the observed empirical patterns can form under the model. Specifically, genomic islands of divergence can form under selection, as shown by the identified islands and their associated traits under divergent selection (Table S3). However, while simulations show that islands are more likely to form at selected loci, we also show that islands can form under neutral processes, and thus the simulations do not exclusively suggest selection as the explanation for our empirically observed genomic divergence. With an increased divergence timeframe, particularly in the absence of gene flow as simulated here, alternative alleles and genotypic combinations are likely to become more strongly divergent or fix as a consequence of drift.

Signatures of reduced Tajima's *D* and nucleotide diversity (in the tawny and Berthelot's pipit) were found in all the strongly divergent genomic regions identified across the speciation event between the tawny and Berthelot's pipit (Fay & Wu, 2000), indicative of selective sweeps after the species split. Another explanation for this pattern could be that Berthelot's pipit genomes have accumulated a higher number of slightly deleterious mutations through mainland to island colonisation compared to the tawny pipit (Leroy et al., 2021). Highly divergent genomic regions between the Canary Islands and Selvagens also consistently exhibited a strong decrease in Tajima's *D* within the Selvagens, commonly associated with large regions of low diversity or ROH. Negative Tajima's *D* can result from either recent selective sweeps or population expansion following a bottleneck (Tajima, 1989). In the case of the Selvagens pipit population we have strong evidence of a very small contemporary N_e with considerable inbreeding (Martin et al., 2023). Consequently, it seems likely that the genomic islands of divergence observed result from selective sweeps, but we cannot rule out the impact of accumulation of deleterious mutations. Such patterns of selection have previously been reported in other wild populations despite small N_e (de Jong et al., 2020) and short divergence timescales (Walsh et al., 2019). This pattern was not observed between other archipelago populations of Berthelot's pipit (Table S3). Genomic islands of divergence between the Canary Islands and Madeiran archipelago population were associated with increased Tajima's *D* values in Madeira which may result from balancing selection or sudden population contractions. Windows with corresponding moderate levels of genetic

diversity may indicate balancing selection is driving divergence within these regions (Lindtke et al., 2017; Tetteh et al., 2009), as inbreeding effects due to a population contraction would be expected to result in low diversity regions.

We detected a potential ~2Mb 'plateau' of divergence between the tawny and Berthelot's pipit, which is conserved across the Berthelot's pipit range (Figure S4). Despite the absence of gene flow between the two sister species it is still surprising to detect such a broad peak of elevated divergence with low genetic diversity in both populations, as this suggests a long-term absence of recombination in this region. Large-scale valleys of recombination commonly coincide with centromeres (Burri et al., 2015; Carneiro et al., 2009), although it is possible that this is as a result of an inversion previous to the speciation event (Kirkpatrick, 2010; Rieseberg, 2001; Sanchez-Donoso et al., 2022). However, lack of recombination maps for Berthelot's pipit or closely related species means we are unable to test this hypothesis here, although long-read sequencing (Shao et al., 2018) in combination with haplotype phasing (Alachiotis et al., 2012; Ferrer-Admetlla et al., 2014) could be utilised to explore this in future studies. The broad 1A divergence peak identified here mapped to a gene dense region with 17 genes, many of which are associated with craniofacial/bone development, migration strategies or the immune response (Table S2). For example, *CMAS*, *TLL12* and *YBX3* are all associated with viral defence (Carette et al., 2009; Ju et al., 2017; Qin et al., 2020); *TSPO*, *GEC1* and *STYK1* are associated with cellular autophagy or apoptosis (Chakrama et al., 2010; Veenman et al., 2007; Zhou et al., 2020); *KCNJ8* increases wound healing (Zhang & Bei, 2015); *ABCC9* is associated with craniofacial defects in vertebrates (Czeschik et al., 2013; Harakalova et al., 2012) and *SCUBE1* modulates BMP signalling during craniofacial development (Tu et al., 2008; Xavier et al., 2009) and is linked to head morphology adaptation in sticklebacks (Hohenlohe et al., 2010). Inversion haplotypes, and potentially the divergence peak on chromosome 1A, have been found to harbour supergene complexes and may be key facilitators for local adaptation and speciation (Kirkpatrick & Barton, 2006; Taylor & Campagna, 2016; for example, Christmas et al., 2019; Küpper et al., 2016; Porubsky et al., 2020; Roesti et al., 2015; Tuttle et al., 2016).

We identified positively selected candidate genes in Berthelot's pipit within regions exhibiting elevated differentiation that could be associated with transition from migration strategies in the tawny pipit to residency and limited dispersal in Berthelot's pipit (Table S2). For example, two genes within the broad divergence peak on chromosome 1A have previously been associated with migratory to residency transition between continental and island populations of the European blackcap (*Sylvia atricapilla*; Delmore et al., 2020); *GYS2*, which is transcriptionally activated by *CLOCK* which regulates circadian rhythms of hepatic glycogen synthesis (Doi et al., 2010). Circadian rhythms synchronise circannual (seasonal) clocks, which can regulate migratory timing and distance (Gwinner, 1996). The second gene is *TLL12*, but the function it plays in regulation of these divergent strategies is poorly understood. A further gene in this broad divergence region, *TSPO*, is associated with metabolism through regulating mitochondrial function and glycolysis

(Veenman et al., 2007; Yao et al., 2020). We also detected another gene, *PDE1C*, in a different divergence peak on chromosome 2, associated with metabolic regulation (Lugnier, 2011) that may also be involved in this behavioural transition in the Berthelot's pipit.

Pathogen prevalence varies substantially (and consistently) between different Berthelot's pipit populations, and between those and the tawny pipit, providing a strong environmental selection gradient (Armstrong et al., 2018; Illera et al., 2008; Spurgin et al., 2012). Studies of Berthelot's pipit have previously identified pathogen-mediated selection associated with key immune genes including Toll-like receptor 4 (*TLR4*) (Armstrong et al., 2019; Gonzalez-Quevedo et al., 2015) and MHC class I genes (Gonzalez-Quevedo et al., 2016; Spurgin et al., 2011) within islands and across archipelagos (Armstrong et al., 2018). In the present study, we identified various genes in highly diverged regions that could be involved with host-pathogen evolution including: *ITGA9*, *SEN7* and *NFKB* which are associated with the innate immune response (Audard et al., 2012; Cui et al., 2017; Sigurðarson, 2020); and *WAC* and *GEC1* which interact in the same pathway to regulate autophagy (Joachim et al., 2015). We did not, however, detect high levels of divergence in regions containing genes associated with pathogen response in comparisons involving the most recently founded Berthelot's pipit archipelago, the Selvagens. This is to be expected given the lack of known pathogens in this very isolated archipelago (Spurgin et al., 2012).

Variation in bill morphology across the species range has been largely attributed to founder effects across the archipelagos (Armstrong et al., 2018; Spurgin et al., 2014). However, our findings suggest that selection may also play a role in shaping craniofacial development through each colonisation event (Tables S2 and S3). Regions of the genome putatively under selection included: *CIT*, *NRXN1*, *PRDM10* and *WAC*, all associated with craniofacial defects in vertebrates (De Santo et al., 2015; Park & Kim, 2010; Shaheen et al., 2016; Zahir et al., 2008); *ABI3BP*, associated with bone development (Zhang et al., 2014); and *SOC7* which interacts with growth factors (Elliott & Johnston, 2004). These genes may underlie variation in bill morphology that enhances exploitation of differing food resources across the range. Again, studies on foraging behaviour and trophic ecology are now needed to unravel the causes explaining such bill variation across archipelagos.

Our findings provide evidence of repeated divergence – potentially as a result of selection – for genes previously found to be involved in head/bill and body size, eye development, metabolism, wound healing and immune defence, across different temporal scales in Berthelot's pipit. However, as genomic islands of divergence rarely occurred at the same location across inter-population comparisons, it is possible that different genes with similar phenotypic effects may be responding to selection in different locations. Parallel adaptation for traits through different candidate genes has been reported across vertebrates (Langin et al., 2015; Milner et al., 1999; Walsh et al., 2019). Overall, our findings add to growing evidence of immunity and head/bill and body size as ecologically important traits for adaptive divergence and speciation (immune; Davies et al., 2021; Hughes & Yeager, 1998; Jarvi et al., 2001, and

bill, Badyaev et al., 2008; Bosse et al., 2017; Dussex et al., 2021; Grant, 1968; Lundregan et al., 2018). Small numbers of individuals sampled per population, as used here, may lead to unrepresentative calculation of population-based statistics, especially between the Tawny pipit and Berthelot's pipit. We therefore caution readers against reliance upon the validity of specific candidate SNPs identified in this study (Tables S2–S4).

4.4 | Genomic valleys of divergence through speciation

Between the tawny and Berthelot's pipit, we detected very few low-divergence genomic regions despite speciation in allopatry; similar to observations of diverging *Ficedula* flycatcher species by Burri et al. (2015). Genomic valleys of divergence may be associated with reduced F_{ST} and nucleotide diversity if they are due to parallel selection (Roesti et al., 2014). Alternatively, they may be associated with reduced F_{ST} and moderate nucleotide diversity if they result from balancing selection (Hohenlohe et al., 2010). In our case, we have limited evidence to confirm either of these possibilities. If the genomic regions we have identified are under parallel/balancing selection in the tawny and Berthelot's pipit, this highlights the importance of a further set of genes associated with body and head size and immune response (Table S4). Interestingly, these candidate genes include *NFX1*, a transcriptional repressor of MHC class II genes (Gewin et al., 2004; Strominger et al., 1994), a gene family known to be under balancing selection in many vertebrates (Meyer & Thomson, 2001; Savage et al., 2020), including avian species (Alcaide et al., 2008; Brouwer et al., 2010; Ekblom et al., 2010). Evidence of the long-term retention of MHC alleles is a well-known phenomenon (Klein, 1987), and balancing selection over similar evolutionary timescales has been reported across populations of many taxa (Bryja et al., 2007; Evans et al., 2010; Herdegen-Radwan et al., 2021; Richardson & Westerdahl, 2003). In Berthelot's pipit, island colonisation is initially associated with reduced MHC diversity but there is evidence of rapid in situ generation of diversity via gene conversion (Spurgin et al., 2011).

5 | CONCLUSIONS

Island colonisation can dramatically influence genetic diversity within and divergence between populations through a combination of evolutionary neutral (i.e. drift) and selective forces. Disentangling the contributions of such processes is particularly complex in cases where there has been strong and/or sequential founder bottlenecks. Our results offer a population-level view of the divergence history of Berthelot's pipit along the speciation continuum, through differing divergence timeframes across three north Atlantic archipelagos. We observed different sets of strongly divergent loci through colonisation of new islands, potentially due to the stochastic nature of population founding. Despite these unique patterns of divergence across

archipelago populations, strongly divergent loci were putatively associated with the same biologically important traits (i.e. pathogen defence, head/body size), probably reflecting parallel adaptation across islands. Future studies encompassing multiple populations such as ours, including more detailed phenotyping and association mapping for a greater number of individuals, may be able to further clarify the relative contribution of genetic drift and selection in shaping the genomic landscape of divergence in wild populations.

AUTHOR CONTRIBUTIONS

Claudia A. Martin, David S. Richardson and Lewis G. Spurgin designed the study, with input from Eleanor C. Sheppard and Alex Suh. Fieldwork was undertaken and permit permissions obtained by Lewis G. Spurgin, David S. Richardson and Juan Carlos Illera. Claudia A. Martin conducted the molecular laboratory work, developed the genetic data sets and performed data analysis, with critical input from Lewis G. Spurgin. Hisham A. A. Ali led the individual-based modelling with input from Claudia A. Martin. Claudia A. Martin drafted the manuscript with assistance from David S. Richardson, and all authors read and contributed to the draft and approved submission of this manuscript.

ACKNOWLEDGEMENTS

We are grateful to Matthew Clark and Lawrence Percival-Alwyn for assistance generating the pipit reference assembly, Maria-Elena Mannarelli for assistance with laboratory work and Martin Taylor and Mike Ritchie for their constructive feedback on this manuscript. Permissions to sample Berthelot's pipits were kindly provided by the Regional Governments of the Canary Islands (Ref.: 2019/5555), the Cabildo de Lanzarote (Ref.: 101/2019), the Cabildo de Tenerife (Ref.: 2019-01740) and Madeira (Ref.: 06/IFCN/2020). We are also grateful to the Governments of the Canary Islands and Madeira for providing accommodation, and the Portuguese Navy for transport to Selvagem Grande. Bioinformatics analyses were carried out on the High-Performance Computing Cluster supported by the Research and Specialist Computing Support service at UEA. This work was supported by Natural Environment Research Council (NERC) studentships awarded to CAM through the EnvEAST DTP (NE/L002582/1) and ECS through the ARIES DTP (NE/S007334/1). Genome sequencing was funded through a Norwich Research Park Science Links Seed Fund. JCI was funded by a research grant from the Spanish Ministry of Science, Innovation and Universities, and the European Regional Development Fund (PGC2018-097575-B-I00) and by a regional GRUPIN grant from the Regional Government of Asturias (AYUD/2021/51261).

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

All data and code used to perform the data analysis within this manuscript are openly available on the Dryad Digital Repository: <https://doi.org/10.5061/dryad.1g1jwsv4b> (Martin et al., 2024).

ORCID

Claudia A. Martin  <https://orcid.org/0000-0003-2645-0790>

Eleanor C. Sheppard  <https://orcid.org/0000-0002-4986-7268>

Hisham A. A. Ali  <https://orcid.org/0000-0001-7467-5199>

Alexander Suh  <https://orcid.org/0000-0002-8979-9992>

Lewis G. Spurgin  <https://orcid.org/0000-0002-0874-9281>

David S. Richardson  <https://orcid.org/0000-0001-7226-9074>

REFERENCES

- Alachiotis, N., Stamatakis, A., & Pavlidis, P. (2012). OmegaPlus: A scalable tool for rapid detection of selective sweeps in whole-genome data-sets. *Bioinformatics*, 28(17), 2274–2275. <https://doi.org/10.1093/bioinformatics/bts419>
- Alcaide, M., Edwards, S. V., Negro, J. J., Serrano, D., & Tella, J. L. (2008). Extensive polymorphism and geographical variation at a positively selected MHC class II B gene of the lesser kestrel (*Falco naumanni*). *Molecular Ecology*, 17(11), 2652–2665. <https://doi.org/10.1111/j.1365-294X.2008.03791.x>
- Alef, T., Torres, S., Hausser, I., Metze, D., Türsen, Ü., Lestringant, G. G., & Hennies, H. C. (2009). Ichthyosis, follicular atrophoderma, and hypotrichosis caused by mutations in ST14 is associated with impaired profilaggrin processing. *Journal of Investigative Dermatology*, 129(4), 862–869. <https://doi.org/10.1038/jid.2008.311>
- Ali, H. A. A., Coulson, T., Clegg, S. M., & Quilodrán, C. S. (2023). The effect of divergent parallel selection on genomic landscape of divergence. *Molecular Ecology*, 33(3), 1–16. <https://doi.org/10.1111/mec.17225>
- Alleaume-Benharira, M., Pen, I. R., & Ronce, O. (2006). Geographical patterns of adaptation within a species' range: Interactions between drift and gene flow. *Journal of Evolutionary Biology*, 19(1), 203–215. <https://doi.org/10.1111/j.1420-9101.2005.00976.x>
- Arctander, P., Folmer, O., & Fjeldsa, J. (1996). The phylogenetic relationships of Berthelot's pipit *Anthus berthelotii* illustrated by DNA sequence data, with remarks on the genetic distance between rock and water pipits *Anthus spinoletta*. *Ibis*, 138(2), 263–272. <https://doi.org/10.1111/j.1474-919X.1996.tb04338.x>
- Armstrong, C., Davies, R. G., González-Quevedo, C., Dunne, M., Spurgin, L. G., & Richardson, D. S. (2019). Adaptive landscape genetics and malaria across divergent Island bird populations. *Ecology and Evolution*, 9(22), 1–21. <https://doi.org/10.1002/ece3.5700>
- Armstrong, C., Richardson, D. S., Hipperson, H., Horsburgh, G. J., Küpper, C., Percival-Alwyn, L., Clark, M., Burke, T., & Spurgin, L. G. (2018). Genomic associations with bill length and disease reveal drift and selection across Island bird populations. *Evolution Letters*, 2(1), 22–36. <https://doi.org/10.1002/evl3.38>
- Audard, V., Pawlak, A., Candelier, M., Lang, P., & Sahali, D. (2012). Upregulation of nuclear factor-related kappa b suggests a disorder of transcriptional regulation in minimal change nephrotic syndrome. *PLoS One*, 7(1), 1–9. <https://doi.org/10.1371/journal.pone.0030523>
- Badyaev, A. V., Young, R. L., Oh, K. P., & Addison, C. (2008). Evolution on a local scale: Developmental, functional, and genetic bases of divergence in bill form and associated changes in song structure between adjacent habitats. *Evolution*, 62(8), 1951–1964. <https://doi.org/10.1111/j.1558-5646.2008.00428.x>
- Barton, N. H., & Charlesworth, B. (1984). Genetic revolutions, founder effects, and speciation. *Annual Review of Ecology and Systematics*, 15, 133–164.
- Bay, R. A., & Rugg, K. (2017). Genomic islands of divergence or opportunities for introgression? *Proceedings of the Royal Society B: Biological Sciences*, 284(1850), 20162414. <https://doi.org/10.1098/rspb.2016.2414>

- Bergström, A., Stanton, D. W. G., Taron, U. H., Frantz, L., Sinding, M. H. S., Ersmark, E., Pfrengle, S., Cassatt-Johnstone, M., Lebrasseur, O., Girdland-Flink, L., Fernandes, D. M., Ollivier, M., Speidel, L., Gopalakrishnan, S., Westbury, M. V., Ramos-Madriral, J., Feuerborn, T. R., Reiter, E., Gretzinger, J., ... Skoglund, P. (2022). Grey wolf genomic history reveals a dual ancestry of dogs. *Nature*, 607(7918), 313–320. <https://doi.org/10.1038/s41586-022-04824-9>
- Berry, R. J. (1986). Genetics of insular populations of mammals, with particular reference to differentiation and founder effects in British small mammals. *Biological Journal of the Linnean Society*, 28, 205–230.
- Bosse, M., Spurgin, L. G., Laine, V. N., Cole, E. F., Firth, J. A., Gienapp, P., Gosler, A. G., McMahon, K., Poissant, J., Verhagen, I., Groenen, M. A. M., van Oers, K., Sheldon, B. C., Visser, M. E., & Slate, J. (2017). Recent natural selection causes adaptive evolution of an avian polygenic trait. *Science*, 358(6361), 365–368. <https://doi.org/10.1126/science.aal3298>
- Brouwer, L., Barr, I., Van De Pol, M., Burke, T., Komdeur, J., & Richardson, D. S. (2010). MHC-dependent survival in a wild population: Evidence for hidden genetic benefits gained through extra-pair fertilizations. *Molecular Ecology*, 19(16), 3444–3455. <https://doi.org/10.1111/j.1365-294X.2010.04750.x>
- Bryja, J., Charbonnel, N., Berthier, K., Galan, M., & Cosson, J. F. (2007). Density-related changes in selection pattern for major histocompatibility complex genes in fluctuating populations of voles. *Molecular Ecology*, 16(23), 5084–5097. <https://doi.org/10.1111/j.1365-294X.2007.03584.x>
- Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., Suh, A., Dutoit, L., Bureš, S., Garamszegi, L. Z., Hogner, S., Moreno, J., Qvarnström, A., Ružić, M., Sæther, S. A., Sætre, G. P., Török, J., & Ellegren, H. (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Research*, 25(11), 1656–1665. <https://doi.org/10.1101/gr.196485.115>
- Cabanettes, F., & Klopp, C. (2018). D-GENIES: Dot plot large genomes in an interactive, efficient and simple way. *PeerJ*, 2018(6), e4958. <https://doi.org/10.7717/peerj.4958>
- Calero-Riestra, M., & García, J. T. (2016). Sex-dependent differences in avian malaria prevalence and consequences of infections on nestling growth and adult condition in the Tawny pipit, *Anthus campestris*. *Malaria Journal*, 15(1), 1–11. <https://doi.org/10.1186/s12936-016-1220-y>
- Carette, J. E., Guimaraes, C. P., Varadarajan, M., Park, A. S., Wuethrich, I., Godarova, A., Kotecki, M., Cochran, B. H., Spooner, E., Ploegh, H. L., & Brummelkamp, T. R. (2009). Haploid genetic screens in human cells identify host factors used by pathogens. *Science*, 326(5957), 1231–1235. <https://doi.org/10.1126/science.1178955>
- Carneiro, M., Ferrand, N., & Nachman, M. W. (2009). Recombination and speciation: Loci near centromeres are more differentiated than loci near telomeres between subspecies of the European rabbit (*Oryctolagus cuniculus*). *Genetics*, 181, 593–606. <https://doi.org/10.1534/genetics.108.096826>
- Chakrama, F. Z., Seguin-Py, S., le Grand, J. N., Fraichard, A., Delage-Mourroux, R., Despouy, G., Perez, V., Jouvenot, M., & Boyer-Guittaut, M. (2010). GABARAPL1 (GEC1) associates with autophagic vesicles. *Autophagy*, 6(4), 495–505. <https://doi.org/10.4161/auto.6.4.11819>
- Cho, Y. B., Hong, S., Kang, K. W., Kang, J. H., Lee, S. M., & Seo, Y. J. (2020). Selective and ATP-competitive kinesin KIF18A inhibitor suppresses the replication of influenza A virus. *Journal of Cellular and Molecular Medicine*, 24(10), 5463–5475. <https://doi.org/10.1111/jcmm.15200>
- Choi, J. Y., Purugganan, M., & Stacy, E. A. (2020). Divergent selection and primary gene flow shape incipient speciation of a riparian tree on Hawaii Island. *Molecular Biology and Evolution*, 37(3), 695–710. <https://doi.org/10.1093/molbev/msz259>
- Christmas, M. J., Wallberg, A., Bunikis, I., Olsson, A., Wallerman, O., & Webster, M. T. (2019). Chromosomal inversions associated with environmental adaptation in honeybees. *Molecular Ecology*, 28(6), 1358–1374. <https://doi.org/10.1111/mec.14944>
- Clegg, S. M., Degnan, S. M., Kikkawa, J., Moritz, C., Estoup, A., & Owens, I. P. F. (2002). Genetic consequences of sequential founder events by an Island-colonizing bird. *Proceedings of the National Academy of Sciences*, 99(12), 8127–8132. <http://www.pnas.org/cgi/doi/10.1073/pnas.102583399>
- Cui, Y., Yu, H., Zheng, X., Peng, R., Wang, Q., Zhou, Y., Wang, R., Wang, J., Qu, B., Shen, N., Guo, Q., Liu, X., & Wang, C. (2017). SENP7 potentiates cGAS activation by relieving SUMO-mediated inhibition of cytosolic DNA sensing. *PLoS Pathogens*, 13(1), 1–24. <https://doi.org/10.1371/journal.ppat.1006156>
- Cvijović, I., Good, B. H., & Desai, M. M. (2018). The effect of strong purifying selection on genetic diversity. *Genetics*, 209(4), 1235–1278. <https://doi.org/10.1534/genetics.118.301058>
- Czeschik, J. C., Voigt, C., Goecke, T. O., Lüdecke, H. J., Wagner, N., Kuechler, A., & Wiczorek, D. (2013). Wide clinical variability in conditions with coarse facial features and hypertrichosis caused by mutations in ABCC9. *American Journal of Medical Genetics, Part A*, 161(2), 295–300. <https://doi.org/10.1002/ajmg.a.35735>
- Dai, F., Lin, X., Chang, C., & Feng, X. H. (2009). Nuclear export of Smad2 and Smad3 by RanBP3 facilitates termination of TGF- β signalling. *Developmental Cell*, 16(3), 345–357. <https://doi.org/10.1016/j.devcel.2009.01.022>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Davies, C. S., Taylor, M. I., Hammers, M., Burke, T., Komdeur, J., Dugdale, H. L., & Richardson, D. S. (2021). Contemporary evolution of the innate immune receptor gene TLR3 in an isolated vertebrate population. *Molecular Ecology*, 30(11), 2528–2542. <https://doi.org/10.1111/mec.15914>
- de Jong, M. J., Li, Z., Qin, Y., Quéméré, E., Baker, K., Wang, W., & Hoelzel, A. R. (2020). Demography and adaptation promoting evolutionary transitions in a mammalian genus that diversified during the Pleistocene. *Molecular Ecology*, 29(15), 2777–2792. <https://doi.org/10.1111/mec.15450>
- De Santo, C., D'Aco, K., Araujo, G. C., Shannon, N., DDD Study, Vernon, H., Rahrig, A., Monaghan, K. G., Niu, Z., Vitazka, P., Dodd, J., Tang, S., Manwaring, L., Martir-Negron, A., Schnur, R. E., Juusola, J., Schroeder, A., Pan, V., Helbig, K. L., ... Shinawi, M. (2015). WAC loss-of-function mutations cause a recognisable syndrome characterised by dysmorphic features, developmental delay and hypotonia and recapitulate 10p11.23 microdeletion syndrome. *Journal of Medical Genetics*, 52(11), 754–761. <https://doi.org/10.1136/jmedgenet-2015-103069>
- Delmore, K., Illera, J. C., Pérez-Tris, J., Segelbacher, G., Lugo Ramos, J. S., Durieux, G., Ishigohoka, J., & Liedvogel, M. (2020). The evolutionary history and genomics of European blackcap migration. *eLife*, 9, 1–24. <https://doi.org/10.7554/eLife.54462>
- Delmore, K. E., Lugo Ramos, J. S., Van Doren, B. M., Lundberg, M., Bensch, S., Irwin, D. E., & Liedvogel, M. (2018). Comparative analysis examining patterns of genomic differentiation across multiple episodes of population divergence in birds. *Evolution Letters*, 2(2), 76–87. <https://doi.org/10.1002/evl3.46>
- Doi, R., Oishi, K., & Ishida, N. (2010). CLOCK regulates circadian rhythms of hepatic glycogen synthesis through transcriptional activation of Gys2. *Journal of Biological Chemistry*, 285(29), 22114–22121. <https://doi.org/10.1074/jbc.M110.110361>
- Dussex, N., Kutschera, V. E., Wiberg, R. A. W., Parker, D. J., Hunt, G. R., Gray, R. D., Rutherford, K., Abe, H., Fleischer, R. C., Ritchie, M. G., Rutz, C., Wolf, J. B. W., & Gemmell, N. J. (2021). A genome-wide

- investigation of adaptive signatures in protein-coding genes related to tool behaviour in new Caledonian and Hawaiian crows. *Molecular Ecology*, 30(4), 973–986. <https://doi.org/10.1111/mec.15775>
- Eklom, R., Sæther, S. A., Fiske, P., Kålås, J. A., & Höglund, J. (2010). Balancing selection, sexual selection and geographic structure in MHC genes of great snipe. *Genetica*, 138(4), 453–461. <https://doi.org/10.1007/s10709-008-9335-x>
- Ellegren, H. (2011). Sex-chromosome evolution: Recent progress and the influence of male and female heterogamety. *Nature Reviews Genetics*, 12(3), 157–166. <https://doi.org/10.1038/nrg2948>
- Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., Künstner, A., Mäkinen, H., Nadachowska-Brzyska, K., Qvarnström, A., Uebbing, S., & Wolf, J. B. W. (2012). The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*, 491(7426), 756–760. <https://doi.org/10.1038/nature11584>
- Elliott, J., & Johnston, J. A. (2004). SOCS: Role in inflammation, allergy and homeostasis. *Trends in Immunology*, 25(8), 434–440. <https://doi.org/10.1016/j.it.2004.05.012>
- Estandia, A., Sendell-Price, A. T., Robertson, B. C., & Clegg, S. M. (2023). Standing genetic variation and de novo mutations underlie parallel evolution of island bird phenotypes. *bioRxiv*. <https://doi.org/10.1101/2023.06.29.546893>
- Evans, M. L., Neff, B. D., & Heath, D. D. (2010). MHC genetic structure and divergence across populations of Chinook salmon (*Oncorhynchus tshawytscha*). *Heredity*, 104(5), 449–459. <https://doi.org/10.1038/hdy.2009.121>
- Fay, J. C., & Wu, C. I. (2000). Hitchhiking under positive Darwinian selection. *Genetics*, 155(3), 1405–1413. <https://doi.org/10.1093/genetics/155.3.1405>
- Feder, J. L., Flaxman, S. M., Egan, S. P., Comeault, A. A., & Nosil, P. (2013). Geographic mode of speciation and genomic divergence. *Annual Review of Ecology, Evolution, and Systematics*, 44, 73–97. <https://doi.org/10.1146/annurev-ecolsys-110512-135825>
- Feder, J. L., & Nosil, P. (2010). The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution*, 64(6), 1729–1747. <https://doi.org/10.1111/j.1558-5646.2009.00943.x>
- Ferrer-Admetlla, A., Liang, M., Korneliusen, T., & Nielsen, R. (2014). On detecting incomplete soft or hard selective sweeps using haplotype structure. *Molecular Biology and Evolution*, 31(5), 1275–1291. <https://doi.org/10.1093/molbev/msu077>
- Florencio, M., Patiño, J., Nogué, S., Traveset, A., Borges, P. A. V., Schaefer, H., Amorim, I. R., Arnedo, M., Ávila, S. P., Cardoso, P., de Nascimento, L., Fernández-Palacios, J. M., Gabriel, S. I., Gil, A., Gonçalves, V., Haroun, R., Illera, J. C., López-Darias, M., Martínez, A., ... Santos, A. M. C. (2021). Macaronesia as a fruitful arena for ecology, evolution, and conservation biology. *Frontiers in Ecology and Evolution*, 9, 1–19. <https://doi.org/10.3389/fevo.2021.718169>
- Garcia-del-rey, E., & Cresswell, W. (2007). The breeding biology of the endemic Berthelot's pipit *Anthus berthelotii* in a harsh oceanic Island environment (Tenerife, Canary Islands). *Ostrich – Journal of African Ornithology*, 78(January), 583–589. <https://doi.org/10.2989/OSTRICH.2007.78.3.5.316>
- Gewin, L., Myers, H., Kiyono, T., & Galloway, D. A. (2004). Identification of a novel telomerase repressor that interacts with the human papillomavirus type-16 E6/E6-AP complex. *Genes and Development*, 18(18), 2269–2282. <https://doi.org/10.1101/gad.1214704>
- Gonzalez-Quevedo, C., Davies, R. G., Phillips, K. P., Spurgin, L. G., & Richardson, D. S. (2016). Landscape-scale variation in an anthropogenic factor shapes immune gene variation within a wild population. *Molecular Ecology*, 25(17), 4234–4246. <https://doi.org/10.1111/mec.13759>
- Gonzalez-Quevedo, C., Davies, R. G., & Richardson, D. S. (2014). Predictors of malaria infection in a wild bird population: Landscape-level analyses reveal climatic and anthropogenic factors. *Journal of Animal Ecology*, 83(5), 1091–1102. <https://doi.org/10.1111/1365-2656.12214>
- Gonzalez-Quevedo, C., Spurgin, L. G., Illera, J. C., & Richardson, D. S. (2015). Drift, not selection, shapes toll-like receptor variation among oceanic Island populations. *Molecular Ecology*, 24(23), 5852–5863. <https://doi.org/10.1111/mec.13437>
- Grabherr, M. G., Russell, P., Meyer, M., Mauceli, E., Alföldi, J., di Palma, F., & Lindblad-Toh, K. (2010). Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics*, 26(9), 1145–1151. <https://doi.org/10.1093/bioinformatics/btq102>
- Grant, P. R. (1968). Bill size, body size, and the ecological adaptations of bird species to competitive situations on islands. *Systematic Biology*, 17(3), 319–333. <https://doi.org/10.1093/sysbio/17.3.319>
- Griffiths, R., Double, M. C., Orr, K., & Dawson, R. J. G. (1998). A DNA test to sex most birds. *Molecular Ecology*, 7, 1071–1075.
- Gwinner, E. (1996). Circadian and circannual programmes in avian migration. *The Journal of Experimental Biology*, 199(Pt 1), 39–48. <https://doi.org/10.1242/jeb.199.1.39>
- Han, P., Werber, J., Surana, M., Fleischer, N., & Michaeli, T. (1999). The calcium/calmodulin-dependent phosphodiesterase PDE1C down-regulates glucose-induced insulin secretion. *Journal of Biological Chemistry*, 274(32), 22337–22344. <https://doi.org/10.1074/jbc.274.32.22337>
- Harakalova, M., van Harsse, J. J. T., Terhal, P. A., van Lieshout, S., Duran, K., Renkens, I., Amor, D. J., Wilson, L. C., Kirk, E. P., Turner, C. L. S., Shears, D., Garcia-Minaur, S., Lees, M. M., Ross, A., Venselaar, H., Vriend, G., Takanari, H., Rook, M. B., van der Heyden, M. A. G., ... Cuppen, E. (2012). Dominant missense mutations in ABCC9 cause Cantú syndrome. *Nature Genetics*, 44(7), 793–796. <https://doi.org/10.1038/ng.2324>
- Harrison, R. G. (1991). Molecular changes at speciation. *Annual Review of Ecology and Systematics*, 22(1), 281–308. <https://doi.org/10.1146/annurev.es.22.110191.001433>
- Hartert, E. (1905). *Die Vögel der paläarktischen fauna*. Vol. 1 (Part 3) (pp. 241–384). R. Friedländer und Sohn.
- Hejase, H. A., Salman-Minkov, A., Campagna, L., Hubisz, M. J., Lovette, I. J., Gronau, I., & Siepel, A. (2020). Genomic islands of differentiation in a rapid avian radiation have been driven by recent selective sweeps. *Proceedings of the National Academy of Sciences of the United States of America*, 117(48), 30554–30565. <https://doi.org/10.1073/pnas.2015987117>
- Herdegen-Radwan, M., Phillips, K. P., Babik, W., Mohammed, R. S., & Radwan, J. (2021). Balancing selection versus allele and supertype turnover in MHC class II genes in guppies. *Heredity*, 126(3), 548–560. <https://doi.org/10.1038/s41437-020-00369-7>
- Hofer, T., Foll, M., & Excoffier, L. (2012). Evolutionary forces shaping genomic islands of population differentiation in humans. *BMC Genomics*, 13(1), 1–13. <https://doi.org/10.1186/1471-2164-13-107>
- Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, 6(2), e1000862. <https://doi.org/10.1371/journal.pgen.1000862>
- Hughes, A. L., & Yeager, M. (1998). Natural selection at major histocompatibility complex loci of vertebrates. *Annual Review of Genetics*, 32(1), 415–435. <https://doi.org/10.1146/annurev.genet.32.1.415>
- Illera, J. C., Emerson, B. C., & Richardson, D. S. (2007). Population history of Berthelot's pipit: Colonization, gene flow and morphological divergence in Macaronesia. *Molecular Ecology*, 16(21), 4599–4612. <https://doi.org/10.1111/j.1365-294X.2007.03543.x>
- Illera, J. C., Emerson, B. C., & Richardson, D. S. (2008). Genetic characterization, distribution and prevalence of avian pox and avian malaria in the Berthelot's pipit (*Anthus berthelotii*) in Macaronesia. *Parasitology Research*, 103(6), 1435–1443. <https://link.springer.com/content/pdf/10.1007/2Fs00436-008-1153-7.pdf>
- Illera, J. C., Spurgin, L. G., Rodriguez-Exposito, E., Nogales, M., & Rando, J. C. (2016). What are we learning about speciation and extinction

- from the Canary Islands? *Ardeola*, 63(1), 15–33. <https://doi.org/10.13157/arla.63.1.2016.rp1>
- Irwin, D. E., Milá, B., Toews, D. P. L., Brelsford, A., Kenyon, H. L., Porter, A. N., Gossen, C., Delmore, K. E., Alcaide, M., & Irwin, J. H. (2018). A comparison of genomic islands of differentiation across three young avian species pairs. *Molecular Ecology*, 27(23), 4839–4855. <https://doi.org/10.1111/mec.14858>
- Jarvi, S. I., Atkinson, C. T., & Fleischer, R. C. (2001). Immunogenetics and resistance to avian malaria in Hawaiian honeycreepers (*Drepanidinae*). *Studies in Avian Biology*, 22(22), 254–263.
- Jiang, Y., Jiang, Y., Wang, S., Zhang, Q., & Ding, X. (2019). Optimal sequencing depth design for whole genome re-sequencing in pigs. *BMC Bioinformatics*, 20(1), 556. <https://doi.org/10.1186/s12859-019-019-3164-z>
- Joachim, J., Jefferies, H. B. J., Razi, M., Frith, D., Snijders, A. P., Chakravarty, P., Judith, D., & Tooe, S. A. (2015). Activation of ULK kinase and autophagy by GABARAP trafficking from the centrosome is regulated by WAC and GM130. *Molecular Cell*, 60(6), 899–913. <https://doi.org/10.1016/j.molcel.2015.11.018>
- Ju, L.-G., Zhu, Y., Lei, P. J., Yan, D., Zhu, K., Wang, X., Li, Q. L., Li, X. J., Chen, J. W., Li, L. Y., & Wu, M. (2017). TTL12 inhibits the activation of cellular antiviral signaling through interaction with VISA/MAVS. *The Journal of Immunology*, 198(3), 1274–1284. <https://doi.org/10.4049/jimmunol.1601194>
- Kimura, M. (1991). The neutral theory of molecular evolution: A review of recent evidence. *The Japanese Journal of Genetics*, 66(4), 367–386. <https://doi.org/10.1266/jjg.66.367>
- Kirkpatrick, M. (2010). How and why chromosome inversions evolve. *PLoS Biology*, 8(9), e1000501. <https://doi.org/10.1371/journal.pbio.1000501>
- Kirkpatrick, M., & Barton, N. (2006). Chromosome inversions, local adaptation and speciation. *Genetics*, 173(1), 419–434. <https://doi.org/10.1534/genetics.105.047985>
- Kirkpatrick, M., & Ravigné, V. (2002). Speciation by natural and sexual selection: Models and experiments. *American Naturalist*, 159(3 Suppl), S22–S35. <https://doi.org/10.2307/3078919>
- Klein, J. (1987). Origin of major histocompatibility complex polymorphism: The trans-species hypothesis. *Human Immunology*, 19(3), 155–162.
- Küpper, C., Stocks, M., Risse, J. E., dos Remedios, N., Farrell, L. L., McRae, S. B., Morgan, T. C., Karlionova, N., Pinchuk, P., Verkuil, Y. I., Kitaysky, A. S., Wingfield, J. C., Piersma, T., Zeng, K., Slate, J., Blaxter, M., Lank, D. B., & Burke, T. (2016). A supergene determines highly divergent male reproductive morphs in the ruff. *Nature Genetics*, 48(1), 79–83. <https://doi.org/10.1038/ng.3443>
- Lamichanay, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., Promerová, M., Rubin, C. J., Wang, C., Zamani, N., Grant, B. R., Grant, P. R., Webster, M. T., & Andersson, L. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, 518(7539), 371–375. <https://doi.org/10.1038/nature14181>
- Langin, K. M., Sillett, T. S., Funk, W. C., Morrison, S. A., Desrosiers, M. A., & Ghalambor, C. K. (2015). Islands within an Island: Repeated adaptive divergence in a single population. *Evolution*, 69(3), 653–665. <https://doi.org/10.1111/evo.12610>
- Le Moan, A., Gaggiotti, O., Henriques, R., Martinez, P., Bekkevold, D., & Hemmer-Hansen, J. (2019). Beyond parallel evolution: When several species colonize the same environmental gradient. *bioRxiv*. <https://doi.org/10.1101/662569>
- Leroy, T., Rousselle, M., Tilak, M. K., Caizergues, A. E., Scornavacca, C., Recuerda, M., Fuchs, J., Illera, J. C., de Swardt, D. H., Blanco, G., Thébaud, C., Milá, B., & Nabholz, B. (2021). Island songbirds as windows into evolution in small populations. *Current Biology*, 31, 1–8. <https://doi.org/10.1016/j.cub.2020.12.040>
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*, 1–3.
- Li, H. (2014). Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics*, 30(20), 2843–2851. <https://doi.org/10.1093/bioinformatics/btu356>
- Li, J. W., Yeung, C. K., Tsai, P. W., Lin, R. C., Yeh, C. F., Yao, C. T., Han, L., Hung, L. M., Ding, P., Wang, Q., & Li, S. H. (2010). Rejecting strictly allopatric speciation on a continental Island: Prolonged postdivergence gene flow between Taiwan (*Leucodiotroon taewanus*, *Passeriformes Timaliidae*) and Chinese (*L. canorum canorum*) hwa-meis. *Molecular Ecology*, 19(3), 494–507. <https://doi.org/10.1111/j.1365-294X.2009.04494.x>
- Liao, W., Atkinson, C. T., LaPointe, D. A., & Samuel, M. D. (2017). Mitigating future avian malaria threats to Hawaiian forest birds from climate change. *PLoS One*, 12(1), 1–25. <https://doi.org/10.1371/journal.pone.0168880>
- Lindtke, D., Lucek, K., Soria-Carrasco, V., Villoutreix, R., Farkas, T. E., Riesch, R., Dennis, S. R., Gompert, Z., & Nosil, P. (2017). Long-term balancing selection on chromosomal variants associated with crypsis in a stick insect. *Molecular Ecology*, 26(22), 6189–6205. <https://doi.org/10.1111/mec.14280>
- Liu, D., Hunt, M., & Tsai, I. J. (2018). Inferring synteny between genome assemblies: A systematic evaluation. *BMC Bioinformatics*, 19(1), 26. <https://doi.org/10.1186/s12859-018-2026-4>
- Lugnier, C. (2011). PDE inhibitors: A new approach to treat metabolic syndrome? *Current Opinion in Pharmacology*, 11(6), 698–706. <https://doi.org/10.1016/j.coph.2011.09.012>
- Lundregan, S. L., Hagen, I. J., Gohli, J., Niskanen, A. K., Kempainen, P., Ringsby, T. H., Kvalnes, T., Pärn, H., Rønning, B., Holand, H., Ranke, P. S., Båtnes, A. S., Selvik, L. K., Lien, S., Sæther, B. E., Husby, A., & Jensen, H. (2018). Inferences of genetic architecture of bill morphology in house sparrow using a high-density SNP array point to a polygenic basis. *Molecular Ecology*, 27(17), 3498–3514. <https://doi.org/10.1111/mec.14811>
- Martin, C. A., Armstrong, C., Illera, J. C., Emerson, B. C., Richardson, D. S., & Spurgin, L. G. (2021). Genomic variation, population history and adaptation between Island bird populations. *Royal Society Open Science*, 8, 201146.
- Martin, C. A., Sheppard, E. C., Ali, H. A. A., Illera, J. C., Suh, A., Spurgin, L. G., & Richardson, D. S. (2024). *Dryad dataset for "Genomic landscapes of divergence among island bird populations: Evidence of parallel adaptation but at different loci?"*. <https://doi.org/10.5061/dryad.1g1jwsv4b>
- Martin, C. A., Sheppard, E. C., Illera, J. C., Suh, A., Nadachowska-Brzyska, K., Spurgin, L. G., & Richardson, D. S. (2023). Runs of homozygosity reveal past bottlenecks and contemporary inbreeding across diverging populations of an Island colonising bird. *Molecular Ecology*, 32, 1972–1989. <https://doi.org/10.1111/mec.16865>
- Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., Blaxter, M., Manica, A., Mallet, J., & Jiggins, C. D. (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research*, 23, 1817–1828.
- Maynard-Smith, J., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research*, 23, 23–35. <https://doi.org/10.1017/S0016672308009579>
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20, 1297–1303.
- Meier, J. I., Marques, D. A., Wagner, C. E., Excoffier, L., & Seehausen, O. (2018). Genomics of parallel ecological speciation in Lake Victoria cichlids. *Molecular Biology and Evolution*, 35(6), 1489–1506. <https://doi.org/10.1093/molbev/msy051>
- Meyer, D., & Thomson, G. (2001). How selection shapes variation of the human major histocompatibility complex: A review. *Annals of Human Genetics*, 1996(1), 1–26. <https://doi.org/10.1046/j.1469-1809.2001.6510001.x>

- Milner, J. M., Albon, S. D., Illius, A. W., Pemberton, J. M., & Clutton-Brock, T. H. (1999). Repeated selection of morphometric traits in the Soay sheep on St Kilda. *Journal of Animal Ecology*, 68(3), 472–488. <https://doi.org/10.1046/j.1365-2656.1999.00299.x>
- Munch, K., Nam, K., Schierup, M. H., & Mailund, T. (2016). Selective sweeps across twenty millions years of primate evolution. *Molecular Biology and Evolution*, 33(12), 3065–3074. <https://doi.org/10.1093/molbev/msw199>
- Nadeau, N. J., Whibley, A., Jones, R. T., Davey, J. W., Dasmahapatra, K. K., Baxter, S. W., Quail, M. A., Joron, M., ffrench-Constant, R. H., Blaxter, M. L., Mallet, J., & Jiggins, C. D. (2012). Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 367(1587), 343–353. <http://rspb.royalsocietypublishing.org/lookup/doi/10.1098/rspb.2011.0198>
- Nei, M. (2005). Selectionism and neutralism in molecular evolution. *Molecular Biology and Evolution*, 22(12), 2318–2342. <https://doi.org/10.1093/molbev/msi242>
- Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics*, 39(1), 197–218. <https://doi.org/10.1146/annurev.genet.39.073003.112420>
- Noor, M. A. F., & Bennett, S. M. (2009). Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity*, 103(6), 439–444. <https://doi.org/10.1038/hdy.2009.151>
- Nosil, P., & Feder, J. L. (2012). Genomic divergence during speciation: Causes and consequences. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 367(1587), 332–342. <https://doi.org/10.1098/rspb.2011.0263>
- Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, 18(3), 375–402. <https://doi.org/10.1111/j.1365-294X.2008.03946.x>
- O'day, E. M., Idos, G. E., Hill, C., Chen, J. W., & Wagner, G. (2018). Cytidine monophosphate N-acetylneuraminic acid synthetase enhances invasion of human triple-negative breast cancer cells. *Oncotargets and Therapy*, 11, 6827–6838. <https://doi.org/10.2147/OTT.S177639>
- Osmond, M. M., & Coop, G. (2020). Genetic signatures of evolutionary rescue by a selective sweep. *Genetics*, 215(3), 813–829. <https://doi.org/10.1534/genetics.120.303173>
- Park, J. A., & Kim, K. C. (2010). Expression patterns of PRDM10 during mouse embryonic development. *BMB Reports*, 43(1), 29–33. <https://doi.org/10.5483/BMBRep.2010.43.1.029>
- Porubsky, D., Sanders, A. D., Höps, W., Hsieh, P. H., Sulovari, A., Li, R., Mercuri, L., Sorensen, M., Murali, S. C., Gordon, D., Cantsilieris, S., Pollen, A. A., Ventura, M., Antonacci, F., Marschall, T., Korbel, J. O., & Eichler, E. E. (2020). Recurrent inversion toggling and great ape genome evolution. *Nature Genetics*, 52(8), 849–858. <https://doi.org/10.1038/s41588-020-0646-x>
- Prugnolle, F., Manica, A., & Balloux, F. (2005). Geography predicts neutral genetic diversity of human populations. *Current Biology*, 15(5), R159–R160. <http://research.marshfieldclinic>
- Qin, Z., Qu, X., Lei, L., Xu, L., & Pan, Z. (2020). Y-box-binding protein 3 (YBX3) restricts influenza A virus by interacting with viral ribonucleoprotein complex and impairing its function. *Journal of General Virology*, 101(4), 385–398. <https://doi.org/10.1099/jgv.0.001390>
- Quilodrán, C. S., Ruegg, K., Sendell-Price, A. T., Anderson, E. C., Coulson, T., & Clegg, S. M. (2020). The multiple population genetic and demographic routes to islands of genomic divergence. *Methods in Ecology and Evolution*, 11(1), 6–21. <https://doi.org/10.1111/2041-210X.13324>
- Räsänen, K., & Hendry, A. P. (2008). Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. *Ecology Letters*, 11(6), 624–636. <https://doi.org/10.1111/j.1461-0248.2008.01176.x>
- Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., Noor, M. A. F., Mehlig, B., & Westram, A. M. (2017). Interpreting the genomic landscape of speciation: A road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, 30(8), 1450–1477. <https://doi.org/10.1111/jeb.13047>
- Ravinet, M., Yoshida, K., Shigenobu, S., Toyoda, A., Fujiyama, A., & Kitano, J. (2017). The genomic landscape at a late stage of stickleback speciation: High genomic divergence interspersed by small localized regions of introgression. <https://doi.org/10.1101/190629>
- Renaut, S., Grassa, C. J., Yeaman, S., Moyers, B. T., Lai, Z., Kane, N. C., Bowers, J. E., Burke, J. M., & Rieseberg, L. H. (2013). Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature Communications*, 4, 1–8. <https://doi.org/10.1038/ncomms2833>
- Richardson, D. S., Jury, F. L., Blaakmeer, K., Komdeur, J., & Burke, T. (2001). Parentage assignment and extra-group paternity in a cooperative breeder: The Seychelles warbler (*Acrocephalus sechellensis*). *Molecular Ecology*, 10, 2263–2273.
- Richardson, D. S., & Westerdahl, H. (2003). MHC diversity in two *Acrocephalus* species: The outbred great reed warbler and the inbred Seychelles warbler. *Molecular Ecology*, 12(12), 3523–3529. <https://doi.org/10.1046/j.1365-294X.2003.02005.x>
- Rieseberg, L. H. (2001). Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution*, 16(7), 351–358.
- Roesti, M., Gavrillets, S., Hendry, A. P., Salzburger, W., & Berner, D. (2014). The genomic signature of parallel adaptation from shared genetic variation. *Molecular Ecology*, 23(16), 3944–3956. <https://doi.org/10.1111/mec.12720>
- Roesti, M., Hendry, A. P., Salzburger, W., & Berner, D. (2012). Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Molecular Ecology*, 21(12), 2852–2862. <https://doi.org/10.1111/j.1365-294X.2012.05509.x>
- Roesti, M., Kueng, B., Moser, D., & Berner, D. (2015). The genomics of ecological vicariance in threespine stickleback fish. *Nature Communications*, 6, 8767. <https://doi.org/10.1038/ncomms9767>
- RStudio Team. (2016). RStudio: Integrated development for R. RStudio, PBC. <http://www.rstudio.com/>
- Ruegg, K., Anderson, E. C., Boone, J., Pouls, J., & Smith, T. B. (2014). A role for migration-linked genes and genomic islands in divergence of a songbird. *Molecular Ecology*, 23(19), 4757–4769. <https://doi.org/10.1111/mec.12842>
- Sanchez-Donoso, I., Ravagni, S., Rodríguez-Teijeiro, J. D., Christmas, M. J., Huang, Y., Maldonado-Linares, A., Puigcerver, M., Jiménez-Blasco, I., Andrade, P., Gonçalves, D., Friis, G., Roig, I., Webster, M. T., Leonard, J. A., & Vilà, C. (2022). Massive genome inversions drives coexistence of divergent morphs in common quails. *Current Biology*, 32(2), 462–469. <https://doi.org/10.1016/j.cub.2021.11.019>
- Savage, A. E., Gratwicke, B., Hope, K., Bronikowski, E., & Fleischer, R. C. (2020). Sustained immune activation is associated with susceptibility to the amphibian chytrid fungus. *Molecular Ecology*, 29, 2889–2903. <https://doi.org/10.1111/mec.15533>
- Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G. P., Bank, C., Brännström, Å., Brelsford, A., Clarkson, C. S., Eroukmanoff, F., Feder, J. L., Fischer, M. C., Foote, A. D., Franchini, P., Jiggins, C. D., Jones, F. C., ... Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics*, 15(3), 176–192. <https://doi.org/10.1038/nrg3644>
- Sendell-Price, A. T., Ruegg, K. C., Robertson, B. C., & Clegg, S. M. (2021). An Island-hopping bird reveals how founder events shape genome-wide divergence. *Molecular Ecology*, 30, 2495–2510. <https://doi.org/10.1111/mec.15898>
- Sendell-Price, A. T., Ruegg, K. C., Anderson, E. C., Quilodrán, Van Doren, B. M., Underwood, V. L., Coulson, T., & Clegg, S. M. (2020). The genomic landscape of divergence across the speciation

- continuum in Island-colonising silvereyes (*Zosterops lateralis*). *bioRxiv*, 21(1), 1–9. <https://doi.org/10.1016/j.solener.2019.02.027>
- Shaheen, R., Hashem, A., Abdel-Salam, G. M. H., Al-Fadhli, F., Ewida, N., & Alkuray, F. S. (2016). Mutations in CIT, encoding citron rho-interacting serine/threonine kinase, cause severe primary microcephaly in humans. *Human Genetics*, 135(10), 1191–1197. <https://doi.org/10.1007/s00439-016-1722-2>
- Shao, H., Ganesamoorthy, D., Duarte, T., Cao, M. D., Hoggart, C. J., & Coin, L. J. M. (2018). nplnv: Accurate detection and genotyping of inversions using long read sub-alignment. *BMC Bioinformatics*, 19(1), 1–13. <https://doi.org/10.1186/s12859-018-2252-9>
- Sheppard, E. C., Martin, C. A., Illera, J. C., Suh, A., Spurgin, L. G., & Richardson, D. S. (2022). Genomic associations with pox virus across divergent island populations in Berthelot's pipit. *Molecular Ecology*, 31(11), 3154–3173. <https://doi.org/10.1111/mec.16461>
- Sigurðarson, A. (2020). Dual RNA-seq analysis of host-pathogen interaction in *Eimeria* infection of chickens.
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. v., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>
- Sims, D., Sudbery, I., Illott, N. E., Heger, A., & Ponting, C. P. (2014). Sequencing depth and coverage: Key considerations in genomic analyses. *Nature Reviews Genetics*, 15(2), 121–132. <https://doi.org/10.1038/nrg3642>
- Soria-Carrasco, V., Gompert, Z., Comeault, A. A., Farkas, T. E., Parchman, T. L., Johnston, J. S., Buerkle, C. A., Feder, J. L., Bast, J., Schwander, T., Egan, S. P., Crespi, B. J., & Nosil, P. (2014). Stick insect genomes reveal natural selection's role in parallel speciation. *Science*, 344(May), 177–190. <https://doi.org/10.4159/harvard.9780674368446.c10>
- Spurgin, L. G., Illera, J. C., Jorgensen, T. H., Dawson, D. A., & Richardson, D. S. (2014). Genetic and phenotypic divergence in an Island bird: Isolation by distance, by colonization or by adaptation? *Molecular Ecology*, 23(5), 1028–1039. <https://doi.org/10.1111/mec.12672>
- Spurgin, L. G., Illera, J. C., Padilla, D. P., & Richardson, D. S. (2012). Biogeographical patterns and co-occurrence of pathogenic infection across Island populations of Berthelot's pipit (*Anthus berthelotii*). *Oecologia*, 168(3), 691–701. <https://doi.org/10.1007/s00442-011-2149-z>
- Spurgin, L. G., & Richardson, D. S. (2010). How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society B: Biological Sciences*, 277(1684), 979–988. <https://doi.org/10.1098/rspb.2009.2084>
- Spurgin, L. G., Van Oosterhout, C., Illera, J. C., Bridgett, S., Gharbi, K., Emerson, B. C., & Richardson, D. S. (2011). Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. *Molecular Ecology*, 20(24), 5213–5225. <https://doi.org/10.1111/j.1365-294X.2011.05367.x>
- Stajich, J. E., & Hahn, M. W. (2005). Disentangling the effects of demography and selection in human history. *Molecular Biology and Evolution*, 22(1), 63–73. <https://doi.org/10.1093/molbev/msh252>
- Stankowski, S., Chase, M. A., Fuiten, A. M., Rodrigues, M. F., Ralph, P. L., & Streisfeld, M. A. (2019). Widespread selection and gene flow shape the genomic landscape during a radiation of monkeyflowers. *PLoS Biology*, 17, 1–31.
- Strominger, J. L., Johns, T., & Song, Z. (1994). A novel cysteine-rich sequence-specific DNA-binding protein interacts with the conserved x-box motif of the human major histocompatibility complex class II genes via a repeated Cys-his domain and functions as a transcriptional repressor. *Journal of Experimental Medicine*, 180(5), 1763–1774.
- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3), 585–595. <https://doi.org/10.1093/genetics/123.3.585>
- Taylor, S., & Campagna, L. (2016). Avian supergenes. *Science*, 351(6272), 446–447.
- Tetteh, K. K. A., Stewart, L. B., Ochola, L. I., Amambua-Ngwa, A., Thomas, A. W., Marsh, K., Weedall, G. D., & Conway, D. J. (2009). Prospective identification of malaria parasite genes under balancing selection. *PLoS One*, 4(5), e5568. <https://doi.org/10.1371/journal.pone.0005568>
- Tomozawa, M., Nunome, M., Suzuki, H., & Ono, H. (2014). Effect of founding events on coat colour polymorphism of *Apodemus speciosus* (Rodentia: Muridae) on the Izu islands. *Biological Journal of the Linnean Society*, 113(2), 522–535. <https://doi.org/10.1111/bj.12348>
- Tu, C. F., Yan, Y. T., Wu, S. Y., Djoko, B., Tsai, M. T., Cheng, C. J., & Yang, R. B. (2008). Domain and functional analysis of a novel platelet-endothelial cell surface protein, SCUBE. *Journal of Biological Chemistry*, 283(18), 12478–12488. <https://doi.org/10.1074/jbc.M705872200>
- Tuttle, E. M., Bergland, A. O., Korody, M. L., Brewer, M. S., Newhouse, D. J., Minx, P., Stager, M., Betuel, A., Cheviron, Z. A., Warren, W. C., Gonser, R. A., & Balakrishnan, C. N. (2016). Divergence and functional degradation of a sex chromosome-like supergene. *Current Biology*, 26(3), 344–350. <https://doi.org/10.1016/j.cub.2015.11.069>
- Tyler, S., & Christie, D. A. (2020). Tawny pipit (*Anthus campestris*), version 1.0. In J. del Hoyo, A. Elliott, J. Sargatal, D. A. Christie, & E. de Juana (Eds.), *Birds of the world*. Cornell Lab of Ornithology. <https://doi.org/10.2173/bow.tawpip1.01>
- Urbanek, K., Sutherland, D. M., Orchard, R. C., Wilen, C. B., Knowlton, J. J., Aravamudan, P., Taylor, G. M., Virgin, H. W., & Dermody, T. S. (2020). Cytidine monophosphate N-acetylneuraminic acid synthetase and solute carrier family 35 member A1 are required for reovirus binding and infection. *Journal of Virology*, 95(2), 1–13. <https://doi.org/10.1128/jvi.01571-20>
- Van Doren, B. M., Campagna, L., Helm, B., Illera, J. C., Lovette, I. J., & Liedvogel, M. (2017). Correlated patterns of genetic diversity and differentiation across an avian family. *Molecular Ecology*, 26(15), 3982–3997. <https://doi.org/10.1111/mec.14083>
- Veenman, L., Papadopoulos, V., & Gavish, M. (2007). Channel-like functions of the 18-kDa translocator protein (TSPO): Regulation of apoptosis and steroidogenesis as part of the host-defense response. *Current Pharmaceutical Design*, 13(23), 2385–2405.
- Via, S. (2012). Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 367(1587), 451–460. <https://doi.org/10.1098/rstb.2011.0260>
- Via, S., & West, J. (2008). The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, 17(19), 4334–4345. <https://doi.org/10.1111/j.1365-294X.2008.03921.x>
- Vijay, N., Ellegren, H., & Wolf, J. B. W. (2017). Genomewide patterns of variation in genetic diversity are shared among populations, species and higher-order taxa. *Molecular Ecology*, 26, 4284–4295. <https://doi.org/10.1111/mec.14195>
- Voelker, G. (1999). Dispersal, vicariance, and clocks: Historical biogeography and speciation in a cosmopolitan passerine genus (*Anthus: Motacillidae*). *Evolution*, 53(5), 1536. <https://doi.org/10.2307/2640899>
- Walsh, J., Benham, P. M., Deane-Coe, P. E., Arcese, P., Butcher, B. G., Chan, Y. L., Cheviron, Z. A., Elphick, C. S., Kovach, A. I., Olsen, B. J., Shriver, W. G., Winder, V. L., & Lovette, I. J. (2019). Genomics of rapid ecological divergence and parallel adaptation in four tidal marsh sparrows. *Evolution Letters*, 3(4), 324–338. <https://doi.org/10.1002/evl3.126>
- Warren, W. C., Clayton, D. F., Ellegren, H., Arnold, A. P., Hillier, L. W., Künstner, A., Searle, S., White, S., Vilella, A. J., Fairley, S., Heger, A., Kong, L., Ponting, C. P., Jarvis, E. D., Mello, C. V., Minx, P., Lovell, P., Velho, T. A. F., Ferris, M., ... Wilson, R. K. (2010). The genome of

- a songbird. *Nature*, 464(7289), 757–762. <https://doi.org/10.1038/nature08819>
- Weisenfeld, N. I., Yin, S., Sharpe, T., Lau, B., Hegarty, R., Holmes, L., Sogoloff, B., Tabbaa, D., Williams, L., Russ, C., Nusbaum, C., Lander, E. S., MacCallum, I., & Jaffe, D. B. (2014). Comprehensive variation discovery in single human genomes. *Nature Genetics*, 46, 1350–1355.
- Weiss-Lehman, C., Tittes, S., Kane, N. C., Hufbauer, R. A., & Melbourne, B. A. (2019). Stochastic processes drive rapid genomic divergence during experimental range expansions. *Proceedings of the Royal Society B: Biological Sciences*, 286(1900), 20190231. <https://doi.org/10.1098/rspb.2019.0231>
- Wolf, J. B. W., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, 18(2), 87–100. <https://doi.org/10.1038/nrg.2016.133>
- Xavier, G. M., Sharpe, P. T., & Cobourne, M. T. (2009). Scube1 is expressed during facial development in the mouse. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 312(5), 518–524. <https://doi.org/10.1002/jez.b.21260>
- Xu, H., Qu, C., Gan, L., Sun, K., Tan, J., Liu, X., Jiang, Z., Tian, W., Liu, W., Zhang, S., Yang, Y., Jiang, L., Zhu, X., & Zhang, L. (2020). Deletion of the Impg2 gene causes the degeneration of rod and cone cells in mice. *Human Molecular Genetics*, 29(10), 1624–1634. <https://doi.org/10.1093/hmg/ddaa062>
- Yan, W., Long, P., Chen, T., Liu, W., Yao, L., Ren, Z., Li, X., Wang, J., Xue, J., Tao, Y., Zhang, L., & Zhang, Z. (2018). A natural occurring mouse model with adgrv1 mutation of usher syndrome 2C and characterization of its recombinant inbred strains. *Cellular Physiology and Biochemistry*, 47(5), 1883–1897. <https://doi.org/10.1159/000491068>
- Yao, R., Pan, R., Shang, C., Li, X., Cheng, J., Xu, J., & Li, Y. (2020). Translocator Protein 18 kDa (TSPO) deficiency inhibits microglial activation and impairs mitochondrial function. *Frontiers in Pharmacology*, 11, 986. <https://doi.org/10.3389/fphar.2020.00986>
- Yousaf, R., Gu, C., Ahmed, Z. M., Khan, S. N., Friedman, T. B., Riazuddin, S., Shears, S. B., & Riazuddin, S. (2018). Mutations in Diphosphoinositol-Pentakisphosphate kinase PPIP5K2 are associated with hearing loss in human and mouse. *PLoS Genetics*, 14(3), 1–20. <https://doi.org/10.1371/journal.pgen.1007297>
- Zahir, F. R., Baross, A., Delaney, A. D., Eyedoux, P., Fernandes, N. D., Pugh, T., Marra, M. A., & Friedman, J. M. (2008). A patient with vertebral, cognitive and behavioural abnormalities and a de novo deletion of NRXN1 α . *Journal of Medical Genetics*, 45(4), 239–243. <https://doi.org/10.1136/jmg.2007.054437>
- Zhang, F., Guo, X., Zhang, Y., Wen, Y., Wang, W., Wang, S., Yang, T., Shen, H., Chen, X., Tian, Q., Tan, L., & Deng, H. W. (2014). Genome-wide copy number variation study and gene expression analysis identify ABI3BP as a susceptibility gene for Kashin-Beck disease. *Human Genetics*, 133(6), 793–799. <https://doi.org/10.1007/s00439-014-1418-4>
- Zhang, W., & Bei, M. (2015). Kcnh2 and Kcnj8 interactively regulate skin wound healing and regeneration. *Wound Repair and Regeneration*, 23(6), 797–806. <https://doi.org/10.1111/wrr.12347>
- Zhou, C., Qian, X., Hu, M., Zhang, R., Liu, N., Huang, Y., Yang, J., Zhang, J., Bai, H., Yang, Y., Wang, Y., Ali, D., Michalak, M., Chen, X. Z., & Tang, J. (2020). STYK1 promotes autophagy through enhancing the assembly of autophagy-specific class III phosphatidylinositol 3-kinase complex I. *Autophagy*, 16(10), 1786–1806. <https://doi.org/10.1080/15548627.2019.1687212>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Martin, C. A., Sheppard, E. C., Ali, H. A. A., Illera, J. C., Suh, A., Spurgin, L. G., & Richardson, D. S. (2024). Genomic landscapes of divergence among island bird populations: Evidence of parallel adaptation but at different loci? *Molecular Ecology*, 33, e17365. <https://doi.org/10.1111/mec.17365>