UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE BIOLOGIA ANIMAL

The genomic impact of deleterious mutations in isolation with migration models

João Guilherme Figueira Lopes Frazão

Mestrado em Biologia Evolutiva e do Desenvolvimento

Dissertação orientada por:
Professor Doutor Vítor Sousa

2019

Resumo

O impacto que a migração entre populações que se estão a adaptar a diferentes ambientes ainda é pouco compreendido a nível genómico. A possibilidade de sequenciar genomas de vários indivíduos de diferentes populações abriu a porta para estudar estes processos e detectar regiões genómicas envolvidas na adaptação local. Dados de espécies com divergência recente sugerem que os padrões de diversidade e diferenciação ao longo do genoma são bastante heterogéneos (e.g., insectos *Timema* e *Heliconius*, peixes *Gasterosteus aculeatus*). Uma explicação para este padrão é que as zonas do genoma com maior divergência (medida através do índice de fixação - F_{ST}) estão sobre selecção divergente e envolvidas na adaptação local ao passo que zonas com pouca diferenciação são neutrais e reflectem o fluxo genético devido à migração. De acordo com esta visão, a heterogeneidade da diversidade e diferenciação ao longo do genoma é sinónimo de divergência com fluxo genético.

No entanto, previsões teóricas recentes indicam que *linked selection* (i.e., o efeito de selecção num alelo afectar os alelos neutrais próximos) podem gerar os mesmos padrões heterogéneos no genoma. Um destes processos é a *background selection* (BGS), que é o efeito que a selecção purificante para remover alelos deletérios tem nos alelos neutrais vizinhos. Este processo reduz a diversidade genética e, como tal, pode levar a um aumento da diferenciação (F_{ST}) entre populações. Estudos anteriores indicam que a magnitude deste processo de BGS é em parte controlada pela taxa de recombinação. Dado que a taxa de recombinação varia ao longo do genoma, espera-se também variação no efeito da BGS, o que pode levar a padrões heterogéneos.

Com mutações deletérias recessivas pode ocorrer um processo de selecção balanceadora denominado "*Associative Overdominance"* (AOD). Este processo ocorre com mutações recessivas com efeito selectivo reduzido. Estas mutações não têm efeitos no fitness quando em heterozigotia e podem aumentar de frequência devido à deriva genética. Logo, em zonas de recombinação reduzida é provável que diferentes genomas acumulem mutações deletérias em diferentes posições. Isto leva a que indivíduos heterozigóticos com dois genomas com mutações deletérias em posições diferentes tenham uma fitness maior que homozigóticos (i.e., *overdominance*). Por sua vez, este tipo de seleção balanceadora afecta os alelos neutrais vizinhos, dado que a AOD leva à manutenção da diversidade genética e de haplótipos com diferentes combinações de mutações deletérias, mantendo a diversidade neutral.

Estes dois processos (AOD e BGS) têm sido usados como uma explicação para a diferente velocidade com que as populações com diferentes efectivos populacionais perdem diversidade. No entanto, são mutuamente exclusivos, a BGS apenas ocorre se houver uma diminuição da frequência das mutações deletérias e a AOD apenas ocorre se houver uma manutenção da diversidade. Estudos teóricos recentes em modelos com uma única população concluíram que o coeficiente de selecção é um dos principais factores envolvidos na transição entre os dois processos.

No entanto, desconhece-se qual o impacto que a BGS e AOD têm quando existe migração entre populações, e como é que estes processos afectam a diferenciação genómica na presença de migração. Esta tese tem como objectivo quantificar o impacto de BGS e AOD nos padrões genómicos de populações que estão a divergir com ou sem fluxo genético. Para tal, usámos um programa de simulação de evolução de populações (SLiM3). Simulámos indivíduos diplóides com genomas compostos por cromossomas de 50 Kb. Considerámos que metade das mutações é deletéria e a outra metade neutral, assumindo o mesmo coeficiente de selecção *s* para todas as mutações deletérias (i.e., fitness para posições homozigóticas é 1-*s*), sendo a fitness de várias posições genómicas multiplicativa. Uma vez

que existem previsões teóricas sobre o impacto da BGS na diversidade de populações isoladas, simulámos uma população isolada com a mesma amostragem e parâmetros que no cenário com duas populações (ver abaixo). Os nossos resultados indicam que, para coeficientes de selecção *s* acima de 0.001 (i.e., coeficiente de selecção relativo $N_e s$>10, existe uma boa concordância entre as previsões e as simulações indicando que o programa de simulações foi implementado correctamente.

Para o caso de duas populações, usámos o modelo *isolation with migration* no qual uma população ancestral se divide em duas populações que passam a evoluir com ou sem migração entre elas. Para que a população ancestral estivesse em equilíbrio selecção-deriva-mutação, nas simulações a população ancestral com um efectivo populacional de 1000 indivíduos evolui isoladamente durante 8000 gerações. As duas populações descendentes têm o mesmo efectivo populacional, e evoluem durante 2000 gerações com ou sem migração entre elas. Ao fim destas 2000 gerações, 20 indivíduos de cada população são amostrados e as mutações neutrais presentes nos seus genomas são analisadas. Para estudar o efeito da migração usámos várias taxas de migração, incluído o valor de zero que corresponde a um cenário sem migração. Variámos outros parâmetros como o coeficiente de selecção *s* (entre 0 e 0.5), o coeficiente de dominância *h* (0.01 e 0.5) e a taxa de recombinação relativa *r* (10, 1 ou 0.1 vezes a taxa de mutação). Para cada combinação de parâmetros fizemos 100 a 200 simulações e estimámos várias estatísticas, como a diversidade genética populacional (através do número de *pairwise differences* - π), a diferenciação entre populações (D_{XY} e F_{ST}), o grau de desvio do Site Frequency Spectrum neutral (através do D de Tajima), o *linkage desequilibrium* (através do R^2) e uma estatística para detectar selecção balanceadora (β).

Os nossos resultados indicam que, no caso de duas populações sem migração entre elas, as mutações co-dominantes com coeficientes de selecção *s* intermédios ($1 < N_e s < 100$) levam a uma diminuição de aproximadamente 50% no π e de 40% no D_{XY}, e a um aumento de 30% no F_{ST} em relação aos níveis neutrais esperados. Este efeito tem uma maior magnitude com taxas de recombinação menores. Os nossos resultados também indicam que a ocorrência de migração diminui o efeito que a BGS tem na redução da diversidade dentro das populações (π) e na distância genética entre populações (D_{XY}). Estudos teóricos anteriores sugerem que a ocorrência de BGS pode levar a aumentos de F_{ST}. Os nossos resultados confirmam esta hipótese, excepto para o caso com taxas de migração mais elevadas do que $2N_e m$=10, em que o F_{ST} não se desvia do esperado para regiões neutrais.

No que diz respeito aos resultados com mutações recessivas, os padrões gerados dependem do coeficiente de selecção. Com selecção intermédia ($1 < N_e s < 100$) há um aumento do π (200%) e do D_{XY} (100%) e uma diminuição do F_{ST} (70%) em relação aos níveis neutrais esperados. Tal como no caso da BGS, estes efeitos têm maior magnitude com uma taxa de recombinação mais baixa. A migração leva qualitativamente aos mesmos padrões, apesar de a migração diminuir a magnitude dos efeitos anteriormente descritos.

Em relação à taxa de recombinação, os nossos resultados indicam que o efeito das mutações deletérias na diversidade genética neutral é visível quando a taxa de recombinação é igual ou inferior à taxa de mutação. Com taxas de recombinação 10 vezes superiores à taxa de mutação, as várias estatísticas indicam que os padrões são semelhantes à espectativa neutral. Uma vez que há uma variação da taxa de recombinação ao longo do genoma, zonas do genoma com elevada taxas de recombinação podem ser usadas como indicadores dos níveis neutral, o que pode ser utilizado como uma referência para testar a ocorrência de BGS ou AOD.

Estudos anteriores em modelos com uma única população indicam que a transição entre AOD e BGS é, em parte, devida ao coeficiente de selecção. Os nossos resultados também mostram que a ocorrência de baixas taxas de migração não leva a uma alteração dos padrões gerados quer pela ocorrência de AOD quer pela ocorrência de BGS. Como tal podemos concluir que populações que experienciam pouca ou nenhuma migração terão assinaturas genómicas semelhantes em zonas com baixas taxas de recombinação, no caso de as mutações deletérias serem o principal factor. Tal hipótese pode ser testada com dados genómicos ao comparar pares de populações com diferentes taxas de migração, dado que esperamos padrões semelhantes em populações com níveis de migração distintos.

Muitos dos testes de selecção divergente existentes são baseados na detecção de *outliers* em *genome scans*, i.e. regiões com estatísticas como F_{ST} ou a diversidade com valores extremos. Os nossos resultados indicam que valores extremos de F_{ST} elevado ou diversidade reduzida pode ser resultado da ocorrência de BGS e não de selecção divergente. Por outro lado, zonas do genoma com F_{ST} reduzido têm sido interpretados como zonas neutrais do genoma afectados pela a migração. Os nossos resultados indicam que mesmo com migração superior a $2N_em>10$, regiões de F_{ST} baixo podem ser gerados por AOD.

O nosso estudo torna claro que a migração não muda os padrões gerados pela ocorrência de *linked selection* e que existe a necessidade de desenvolver métodos de detecção de selecção que tenham em conta a variação na taxa de recombinação ao longo do genoma. Para tal, irá ser necessário no futuro estudar a interacção entre os factores analisados nesta tese e selecção positiva e seleção divergente.

Palavras-chave: Background Selection, Associative Overdominance, Isolamento com migração, Recombinação.

Abstract

The effects of background selection (BGS), i.e. the effect that purifying selection due to removal of deleterious mutations at given site has on linked neutral sites, is well understood in single population models. For recessive deleterious mutations, heterozygotes can have a higher fitness leading to associative overdominance (AOD). Previous studies suggest that BGS may increase genomic differentiation, misleading genomic scans that rely on high differentiation regions to detect genes involved in local adaptation. However, little is known about the interaction of BGS and AOD with gene flow. To characterize the genomic impact of BGS, AOD and migration we used a forward-in-time simulator implemented in the program SLiM 3.2. We considered an isolation with migration model with two populations of constant size. To understand the relative impact of each parameter we used various combinations of migration rates, recombination rates, selective coefficients and dominance coefficients. We find that, in relation to neutral expectations, co-dominant deleterious mutations decrease within population diversity and increase genetic differentiation (F_{ST}) between populations, although with high migration ($2N_em>10$) F_{ST} does not deviate from neutral expectation. Consistent with the effect of AOD, for recessive deleterious mutations we found an increase of neutral diversity for selective coefficients *s* between 0.0001 and 0.1, when recombination rate is lower than the mutation rate. AOD also leads to a decrease in population differentiation (F_{ST}), with higher migration rates decreasing the magnitude of this effect. Thus, BGS and AOD can lead to heterogeneous genomic patterns and bias the detection of divergent selection.

Keywords: Background Selection, Associative Overdominance, Isolation with migration, Hitchhiking.

Table of Contents

List of Figures and Tables

Tables

Figures

Abbreviations

AOD – Associative Overdominance
BGS – Background Selection
DNA - Deoxyribonucleic acid
LD – Linkage Disequilibrium
N_e - Effective population size
SFS - Site Frequency Spectrum
SNP – Single Nucleotide Polymorphism
SSW - selective sweep

1 State of the Art

The recent access to large amounts of genomic data from several individuals from populations adapted to different environments has changed our understanding of the genetic basis of adaptation. This has led to the widespread use of genomic scans as a tool to detect genes under selection (Hoban et al. 2016). One of the statistics used in genome scans is the F_{ST}, which measures the relative genetic differentiation between two populations (Nei 1973). These genome scan tests of differentiation rely on the fact that, if a certain gene or region of the genome is under positive selection in the environment of one population but not in another, or if different alleles are favoured in different environments (divergent selection), there will be an increased differentiation in that gene or region of the genome. It is further assumed that the demographic history of the populations affects the genome-wide patterns and reflect the neutral distribution of F_{ST}, whereas genomic regions with extremely high F_{ST} (outliers) correspond to potential genes under positive or divergent selection. Another of the measures used to detect selection is genetic diversity. These tests are based on the fact that positive selection affects linked neutral variation, a process known as a selective sweep (SSW) (Barton 2000). A selective sweep occurs when a positively selected allele increases in frequency and, due to the linkage between it and neighbouring neutral sites, it will also increase the frequency of linked neutral alleles. Because of recombination, selective sweeps reduce only the diversity of sites closely linked to regions of the genome affected by selection, and diversity levels recover to the neutral expectation as we move away from selected sites. The comparison of the diversity at each region with the global mean value of all genome will allow to detect outliers under putative selection (Hoban et al. 2016; Booker et al. 2017). However, demographic history and neutral processes are also expected to lead to variation along the genome, and hence genome scans rely on statistical tests to detect outliers that model the variation expected due to the inferred demographic history of populations (Beaumont and Nichols 1996; Wolf and Ellegren 2017).

The use of genome scans has revealed high heterogeneity in levels of differentiation and diversity along the genome of many species (Nosil et al. 2009). Such heterogeneity has been predicted by theoretical models that have shown that local adaptation may lead to an increase in genetic differentiation (Tigano and Friesen 2016), especially in populations with migration between them. This is because migrant individuals may carry genes that are maladaptive in the receiving population. The introgression of these maladapted genes into the receiving population will cause the individuals that express them to have a lower fitness, decreasing the probability of these genes to be transmitted to future generations. In contrast, neutral regions can freely flow between populations because they do not impose a fitness cost and, as such, may introgress and recombine freely between populations. This suggests that heterogeneous genomic patterns of differentiation are expected under scenarios of divergent selection in face of gene flow. However, theoretical studies suggest that heterogeneous differentiation along the genome may also be caused by purifying selection due to removal of deleterious mutations acting on linked neutral mutations, a process known as background selection (Charlesworth et al. 1993).

Background selection (BGS) was first described by (Charlesworth et al. 1993) as an explanation for the pattern found by (Begun and Aquadro 1992), that regions with low recombination rates have a reduced genetic diversity when compared with the rest of the genome. The explanation lies in the fact that the removal of deleterious mutations from zones of the genome with low recombination also leads to the removal of neutral mutations in linkage with the removed mutations, reducing neutral diversity. In regions of high recombination, neutral mutations have a lower chance of being lost due to removal of linked deleterious mutations, thus resulting in a positive correlation between recombination rate and genetic diversity (Hudson and Kaplan 1995).

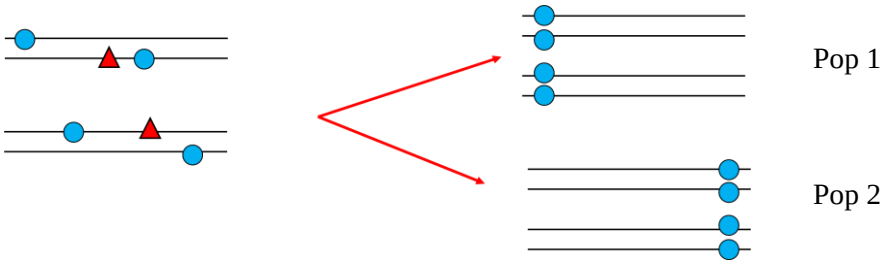

Figure 1.1 ***The effects of background selection on linked neutral mutations.*** *This figure shows four genomes (black lines) that have deleterious mutations (red triangles) that are in linkage with neutral mutations (blue circles). Over time (left side of the figure), deleterious mutations are removed and that causes linked neutral mutations to be removed as well. This leads to a lowered diversity in this region. This will likely increase differentiation, as different populations are likely to loose different mutations. Adapted from (Booker et al. 2017)*

Most theoretical work on BGS has been developed for single populations. Theoretical results showed that the effects of BGS can be approximated by a reduction of the effective population size (N_e) (Hudson and Kaplan 1995). Due to the removal of chromosomes (individuals) that have accumulated deleterious mutations, the dynamics can be described by a population without such individuals and thus with a smaller effective size N_e. The reduction of population size also leads to an increase in the effects of genetic drift and, as such, increased differentiation in allele frequencies between populations. This prediction has been confirmed by some theoretical results that have suggested that BGS may increase F_{ST} in populations with (Zeng and Corcoran 2015) and without (Charlesworth et al. 1997; Nordborg 1997) migration. One of the main factors governing the strength of BGS is recombination rate (Hudson and Kaplan 1995; Nordborg et al. 1996). This is due to the fact that low recombination rates lead to tight linkage between the selected deleterious mutation and the neighbouring neutral mutations. Due to this, the co-occurrence due to linkage of a deleterious and a neutral mutation in the same haplotype will lead the removal of the both mutations, ultimately leading to a greater reduction of diversity. Since recombination rate varies along the genome (Comeron et al. 2012), it is expected that the magnitude of BGS also varies along the genome (Nordborg et al. 1996).

Other phenomena can lead to a statistical association between neutral and deleterious mutations like the effect seen in regions with low recombination rates, including inbreeding and gene conversion (Chen et al. 2007). Gene conversion is a process of DNA repair that is based on the replacement of a track of the genome by the homologous sequence. The interaction that gene conversion (Campos et al. 2017; Pouyet et al. 2018; Campos and Charlesworth 2019) and inbreeding (Charlesworth et al. 1997; Bersabé et al. 2016) have with linked deleterious selection has also been studied from a theoretical perspective. The results from such studies indicate that gene conversion and inbreeding lead to a reduction in neutral diversity that is similar the one expected for the case of BGS (Charlesworth et al. 1997; Bersabé et al. 2016; Campos et al. 2017; Campos and Charlesworth 2019; Gilbert et al. 2019).

Another factor that increases the statistical association between neutral and deleterious mutations is the presence of modifiers of the recombination rate, such as reductions of recombination due to chromosomal inversions (Kirkpatrick and Barton 2006). Inversions have been associated with local adaptation due to the fact that they inhibit crossing-over between chromosomes with and without the inversion (Stevison et al. 2011). Thus, in a scenario of divergence with gene flow, inversions (and regions of low recombination) behave as barriers to gene flow (Sousa and Hey 2013). This leads to a lowered probability of invasion in this region by a maladapted allele, eventually leading to accumulation of mutations under divergent selection in such regions (Feder et al. 2012).

Another aspect that determines the effects of the removal of deleterious mutations on linked neutral variation is the dominance of deleterious mutations. If deleterious mutations are recessive, then it is possible that pseudo-overdominance occurs, a phenomenon where heterozygotic individuals have greater fitness than homozygotic ones due to haplotypes with deleterious mutations in different positions (Gilbert et al. 2019). This is expected to occur when mutations are slightly deleterious and hence can increase in frequency and even fix. To explain this process, assume a scenario where all haplotypes in the population have the same number of recessive deleterious mutations, but in different positions. Since all haplotypes have the same number of recessive mutations their relative fitness is 1.0. However, since deleterious mutations are in different positions in different haplotypes and all deleterious mutations are recessive, then heterozygotes have a higher fitness than homozygotes. This may lead to the maintenance of two or more haplotypes segregating in the population. This is a form of balancing selection that will lead to the persistence of haplotypes that have deleterious mutations in different positions, which are selected against in homozygote state but that are favoured in heterozygotes. This causes the neutral mutations linked to the different haplotypes with deleterious mutations to also be maintained in the population, which increases genetic diversity. This process has been termed associative overdominance (AOD) (Otha 1971). Recent theoretical (Zhao and Charlesworth 2016) and experimental (Schou et al. 2017) results suggest that AOD can lead also to a retardation in the loss of diversity in small populations. Other studies suggested that, in order for AOD to occur, the scaled selective coefficient $N_e s$ should be bellow 100 (Gilbert et al. 2019).

The distribution of dominance and selective coefficients of deleterious mutation is one of the factors that determines if AOD or BGS is more prevalent in the genome. Mutation accumulation and gene deletion studies in *Drosophila melanogaster* (Fry and Nuzhdin 2003) and *Saccharomyces cerevisiae* (Agrawal and Whitlock 2011) have suggested that deleterious mutations are usually recessive. In these organisms, the selective coefficient of deleterious mutations has been found to be between 0.1 and 0.3 for *D. melanogaster* (García-Dorado et al. 1998) and on average 0.045 for *S. cerevisiae* (Agrawal and Whitlock 2011). According to the theoretical results found by (Gilbert et al. 2019) the occurrence of AOD is likely in populations with an effective size smaller than 1,000 for *D. melanogaster* and smaller than 10,000 for *S. cerevisiae.*

Because the occurrence of BGS results from the removal of deleterious mutations and the occurrence of AOD depends on the maintenance of haplotypes with slightly deleterious mutations, these two processes are incompatible. As such, there is a transition between the two, and the conditions that govern the transition between AOD and BGS have been the subject of recent studies. It was found that the selection coefficient and the dominance of the deleterious mutations are probably the main factors that determine the transition from AOD to BGS (Pamilo and Pálsson 1998; Pálsson and Pamilo 1999; Paælsson 2004; Bersabé et al. 2016; Gilbert et al. 2019). However, all of these studies assumed single populations models. Most populations are sub-divided in sub-populations connected by migration. Furthermore, there is a growing interest to understand the conditions under which populations can adapt to different environments with gene flow. However, the role that migration has on the AOD to BGS transition, and how the interaction between these two processes and migration impacts population differentiation at the genomic level is still poorly understood.

Theoretical studies have shown that the transition between the occurrence of BGS and the occurrence of AOD is dependent of the value of the effective selection (the selection coefficient times the population size) and the dominance (Zhao and Charlesworth 2016). However these results were obtained for a single population, and as such, the effects that other parameters may have, especially in models with migration between populations are not yet well understood. Distinguishing the effects that lead to BGS

and AOD and their impact on the genomic patterns is one of the important steps in understanding how to interpret genomic patterns from natural populations. This is required in order to develop methods to detect selection accounting for migration, BGS and AOD.

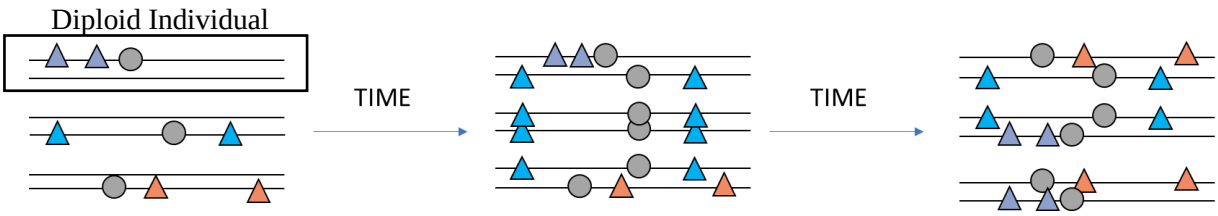

Figure 1.2 ***The effect of Associative Dominance (AOD) on genomic patterns.*** *Each part of the image show three dyploid genomes (black lines) that have different deleterious mutations (blue and orange triangles) that are in linkage with neutral mutations (blue circles) First, slightly recessive deleterious mutations appear in each genome and rise in frequency due to their reduced fitness costs and the effects of drift. After some time (middle image), the deleterious mutations can become frequent enough that all chromosomes have deleterious mutations. As such, the individuals that are homozygotic for the deleterious mutations (individual in the centre) have lower fitness than heterozygotic ones (individuals in the edges). Thus, heterozygotes are favoured, which is a type of balancing selection that maintains the various haplotypes. The linked neutral variation is also maintained.*

2 Objectives

The main goal of this thesis is to characterize the interaction between migration and various forms of linked purifying selection due to removal of deleterious mutations, aiming to answer the following questions:

- What is the effect of recombination on Associative Overdominance (AOD) and background selection (BGS) on genomic patterns under models of isolation with migration?
- What is the impact of BGS and AOD on genomic differentiation and diversity when there is migration?

To answer these questions we performed a simulation study under a model of population divergence (isolation with migration model with two populations). We used individual-based simulations to evaluate the impact of several evolutionary processes acting at the genome level, including recombination rate, selection coefficient and dominance of deleterious mutations, effective population sizes and migration rates. We quantified the genomic impact of linked selection on diversity and population differentiation in isolation with migration models with varying levels of gene flow, by calculating several summary statistics that are widely used in genome scans. We also investigated how the transition between BGS and AOD depends on the combination of parameters such as recombination rate, selective coefficient, dominance and migration rates.

3 Methods

To understand the interaction between migration and linked selection we simulated populations under various scenarios, studying several combinations of parameter values. To do this we performed forward-in-time individual-based simulations, implemented in the program SLiM 3.2 (Haller and Messer 2018). We considered single and two population models, and several combinations of parameters, described in separate sub-sections below.

3.1 Single population model

Since most theoretical results about BGS were obtained for single population models, to test if SLiM simulations fit theoretical expectations of BGS and to verify that our simulations were implemented in SLiM correctly, we first considered a model with a single population evolving alone for 10^5 generations. The genomic structure, sampling, selective coefficients, dominance and recombination rates were the same as in the scenario of an isolation with migration model with two populations (see below). We then compared the results of these simulations with known theoretical results obtained for a single population (Hudson and Kaplan 1995; Campos et al. 2017).

3.2 Isolation with Migration model

To quantify the effect of migration and deleterious mutations on genomic patterns we considered an isolation with migration model, with an ancestral population splitting into two populations with the same effective size as the ancestral. To ensure that the ancestral population is in mutation-selection-drift equilibrium, the ancestral population (effective size N_e=1000) evolves alone for 8000 generation. Then, it splits into two descending sub-populations (N_e=1000 each), which then evolve for 2000 generations (scaled split time of τ=(time in generations)/($2N_e$)=1.0). We assumed that since the split time there is continuous migration, occurring at a constant rate m (corresponding to a scaled migration rate of $2N_em$). To compare cases without migration, we also considered a migration rate of zero.

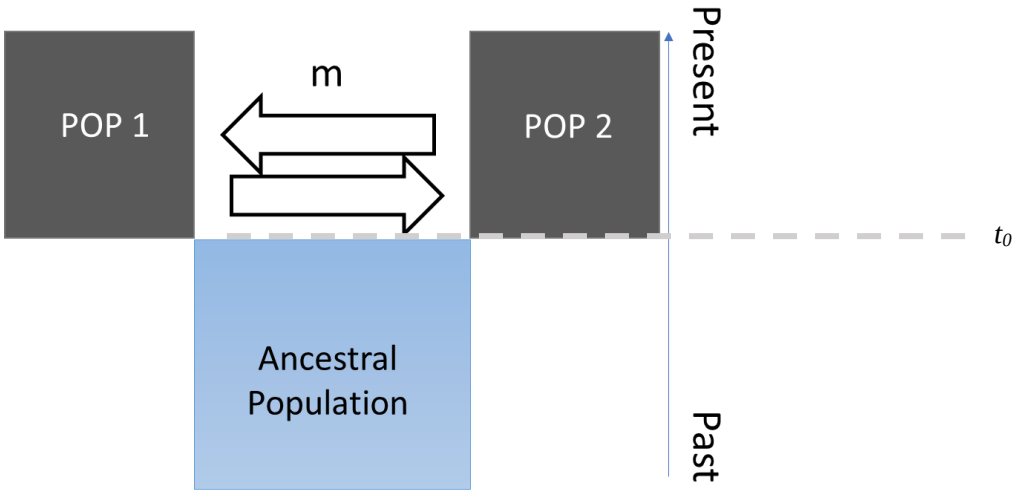

Figure 3.1 ***The isolation with migration model with two populations considered.*** *To ensure that the ancestral population is at mutation-selection-drift equilibrium, we let the ancestral population evolve alone for 8000 generations. We used 8000 generations because this is 8 times larger than the N_e and hence it is enough to attain equilibrium. After this, the ancestral population splits into two populations at time t_0 (2000 generations ago) after which they exchange migrants at a constant rate m (which can be 0 or greater) until the present time (scaled migration rate measured as $2N_em$). N_e was assumed to be the same (1000) in all populations. Note that the scaled divergence time $t_0/(2N_e)$ is 1.0.*

3.3 Replicates and sampling

For each combination of the parameters described in detail below, we performed 100 replicate simulations for the scenario with one population and 200 replicate simulations for the isolation with migration scenario with two populations. At the end of the simulation we sampled the genomes of 20 diploid individuals from each population (i.e., 40 gene copies from each population). Because SLiM allows the modelling of different mutation classes, this allowed to output only the neutral mutations while also simulating the effects of deleterious mutations.

3.4 Genomic structure

In our simulations we assumed that each diploid individual had two copies of a single homologous chromosome with 50Kb. To avoid the possibility that a single site could have neutral and deleterious mutations, we only allowed each site to either have neutral mutations or deleterious mutations. We considered a sequence of alternating neutral and deleterious sites, i.e. the proportion of sites that could potentially have deleterious mutations was 0.50. This choice is in part justified by the structure of a codon. In coding regions, if we consider that non-synonymous mutations are deleterious, then such mutations likely occur when the first or second nucleotides of a codon mutates, since the first two positions determine the amino acid that is codified for most codons. In contrast, the third nucleotide is usually redundant and does not change the amino acid that is codified, resulting in synonymous mutations. Thus, under this simplifying assumption, the proportion of deleterious sites in a coding region would be approximately 2/3. In order to simplify the structure of the simulated genome and computation of summary statistics we assumed that this ratio is 1/2 rather than 2/3. We also assumed that the deleterious and neutral mutation rate was the same, and hence the probability of neutral and deleterious mutations appearing each generation was equal.

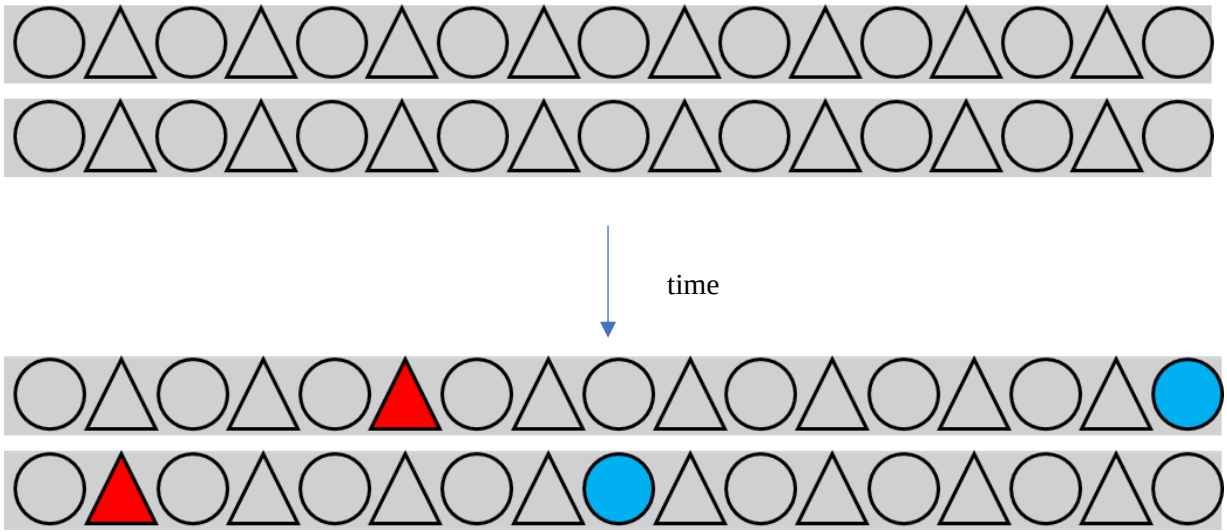

*Figure 3.**2 Genome structure of the diploid individuals in our simulations.** Each figure (triangle or circle) represents a single position in the chromosome where only one type of mutation can occur, triangles represent sites where deleterious mutations can occur and circles represents sites where neutral mutations can occur. In the beginning of the simulation (top image) the genome starts without any mutations. As time goes on, mutations appear in the genome, both deleterious (in red) and neutral (in blue) are assumed to occur at the same rate μ.*

For each site where deleterious mutations could occur, we assumed the following fitness landscape: 1.0 for homozygote for the ancestral allele (i.e., if it had no deleterious mutations), 1-*hs* for heterozygous sites (i.e. with one ancestral and one deleterious allele) and 1-*s* for homozygous sites for the deleterious allele, where *s* is the selection coefficient and *h* is the dominance coefficient of the deleterious mutation considered (*h*=0.5 indicates co-dominant and *h*=0.0 indicates fully recessive). For simplicity, we assumed the same selective and dominance coefficients for all deleterious mutations. The fitness of a

given individual was computed as the product of the fitness for each site, and hence fitness was multiplicative with no epistatic interaction occurring between the different loci.

3.5 Selective coefficients

Because the selective coefficient is thought to be one of the main causes of transition between BGS and AOD, many different values across different scales were used to determine the selective coefficients at which this transition occurs. We assumed that all deleterious mutations have a fixed constant selective coefficient to reduce the complexity of the analysis. We performed simulations varying the selective coefficient values *s* between 0.0001 and 0.5. More specifically we tested the following values: 0.0001, between 0.001 and 0.009 with an increment of 0.001; between 0.01 and 0.09 with an increment of 0.01 and between 0.1 and 0.5 with an increment of 0.1. We also ran simulation with the selective coefficient equal to 0 in order to compare with the known theoretical results of neutral theory and as a way of establishing a baseline with which to compare our results for cases where no theoretical results were available.

3.6 Recombination rate

Because the impact of BGS on diversity is governed by the ratio between recombination and mutation rate (Hudson and Kaplan 1995), the mutation rate was set to a constant (2.5×10^{-7}/site/generation), but the recombination rate was allowed to vary in relation to the mutation rate, taking the values of 0.1, 1 and 10 times the mutation rate (2.5×10^{-8}, 2.5×10^{-7} and 2.5×10^{-6} per site per generation). This corresponds to the probability that a recombination event occurs between two adjacent given sites at each generation, in relation to the probability that there are mutations at those sites. Previous studies show that, in primates the ratio between the recombination rate and the mutation rate varies approximately between 0.1 and 4.0 (Auton et al. 2012) while in insects this ratio is slightly higher than 1 (Terhorst et al. 2017). We used the same ratios of 0.1, 1.0 and 10.0 as used in the latter study (Terhorst et al. 2017) in order to explore a realistic range of values.

3.7 Dominance

Because AOD is expected to only occur if deleterious mutations are recessive, we tested the effect of dominance by using two fixed dominance coefficients (h=0.01 and h=0.5) for each combination of other parameters.

3.8 Migration rate

In the isolation with migration models with two populations we varied the migration rate m (i.e. the probability of migration per individual per generation) between 0.00005 and 0.005. We considered three migration rates, corresponding to scaled migration rates of $2N_e m = 0.1$, 1 and 10 immigrants per generation. The latter was chosen as the highest value because with values of $2N_e m \geq 1$, migration is stronger than genetic drift and hence it counterbalances the loss of diversity and increased differentiation due to genetic drift.

3.9 Effective population size

We considered that the population size of the ancestral and two descending populations was constant and equal in all populations (N_e=1,000 individuals). To understand if the strength of genetic drift affected our results, we also ran simulations with a single population with a larger N_e of 10,000 individuals. Due to time constraints we tested a smaller set of selection coefficients, between the smaller and higher of the two population simulations. In this case, we used the same relative scaled times (i.e., in terms of time/($2N_e$)), both for the split and for reaching equilibrium in the ancestral population, which were set to t_0=20,000 and 80,000 generations, respectively.

Table 8.1 ***Parameters used in the simulations ran with Slim 3.2. – where not applicable.***

Scenario	Mutation rate (μ)	Selective Coefficient (s)	Effective size (N_e)	Migration rate ($2N_em$)	Dominance Coefficient (h)	Recombination rate (r)	Repetitions
Single Population	$2.5x10^{-7}$ /site /generation	Between 0 and 0.1	10,000	-	0.01 and 0.5	0.1, 1 and 10 times the mutation rate μ	100
Single Population	$2.5x10^{-7}$ /site /generation	Between 0 and 0.5	1,000	-	0.01 and 0.5	0.1, 1 and 10 times the mutation rate μ	100
Isolation with Migration	$2.5x10^{-7}$ /site /generation	Between 0 and 0.5	1,000	0, 0.1, 1 and 10 immigrants /generation	0.01 and 0.5	0.1, 1 and 10 times the mutation rate μ	200

Scenario	Effective population size – Parent Population	Effective population size – Derived Populations	Divergence time	Time until divergence
Single Population	1000 Individuals	-	-	-
Isolation with Migration	1000 Individuals	1,000 Individuals	2000 generations	8000 generations

3.10 Summary statistics

We performed the treatment of the SLiM output with the software R (R Core Team 2018).

To measure the nucleotide diversity (π) for each population we calculated the average number of pairwise differences between two sequences sampled in the same population. The following formula was used:

$$\pi = 2\frac{n}{(n-1)}\sum_{i=1}^{L} p_i(1-p_i) \qquad \text{Equation 8.1}$$

where i is the index of each site, p_i is allelic frequency of site i, n is the sample size (number of sequences sampled) and L is the size of the sequence.

In order to quantify the differentiation between populations for the isolation with migration models with two populations we used two complementary statistics: (i) the absolute genetic distance D_{xy}, which measures the mean number of pairwise differences between two sequences taken from two different populations; and (ii) the relative differentiation given by the fixation index (F_{ST}) according to the

estimator of (Bhatia et al. 2013), which measures the "the correlation between randomly drawn alleles from a single population relative to the most recent common ancestral population" (Bhatia et al. 2013). The F_{ST} is a relative measure of differentiation, whereas D_{xy} measures the absolute divergence. For SNPs with two alleles, an F_{ST} of 0.0 for a given locus indicates that the same allele is shared between both populations with exactly the same frequency and an F_{ST} of 1.0 indicates that each population has fixed a different allele. To measure the D_{xy} the following formula was used:

$$D_{xy} = \sum_{i=1}^{L} (p_{i1}(1 - p_{i2})) + (p_{i2}(1 - p_{i1}))$$

Where, $p_{ij}$is allelic frequency of site *i* in population *j*. To measure the F_{ST}, the following formula (equation 10 from (Bhatia et al. 2013)) was used for each site *i*:

$$F_{ST_i} = \frac{(p_{i1} - p_{i2})^2 - \frac{p_{i1}(1 - p_{i1})}{n_1} - \frac{p_{i2}(1 - p_{i2})}{n_2}}{D_{xy}} \qquad \text{Equation 8.3}$$

Where $p_{ij}$is allelic frequency of site *i* in population *j*, and n_j is the sample size of population *j*. After calculating their value, the ratio of the value of these three statistics to the expected values under neutrality was calculated. The expected equilibrium value for diversity π is given by $4N_e\mu L$ (Watterson 1975), where μ is the mutation rate per site per generation and L is the total length of the locus. We investigated the patterns for the deleterious and neutral sites and, given that half of the sites were neutral, we used μ(L/2). The expected values under neutrality for D_{xy} and F_{ST} where obtained from equation 22 and 23 in (Wilkinson-Herbots 2008), and equation 24 of (Zeng and Corcoran 2015), respectively.

To test deviations from the neutral expectations of the site frequency spectrum (SFS), we calculated the Tajima's D for each population (Tajima 1989). This method compares the scaled mutation rate θ ($4N_e\mu$), when inferred from the nucleotide diversity and when inferred from the number of segregating sites (Hamilton 2009). The neutral expectation of Tajima's D is 0, with a positive or a negative value indicating an excess or a lack of diversity given the expected value for the number of segregating sites in the sample.

To measure the effects of BGS, AOD and migration on linkage disequilibrium we calculated the average r^2 for all pairs of SNPs within the 50Kb chromosome for each population (Hill and Robertson 1968). This measure is based on the difference between the observed co-occurrence of two alleles at different loci and their expected co-occurrence given their allelic frequencies assuming free recombination, that is, independent segregation. Higher values of r^2 correspond to higher linkage disequilibrium (Hill and Robertson 1968).

For recessive slightly deleterious recessive mutations AOD can occur, and this can be seen as a form of balancing selection. This is due to the fact that heterozygotes have a higher fitness and hence selection maintains haplotypes with deleterious mutation in different sites. To detect regions under AOD in the genome we computed β, the summary statistic of (Siewert and Voight 2017) which was proposed to detect balancing selection. We developed a script to implement a function that computes this statistic in the R language. This test is based on the fact that in a region of the genome under balancing selection, SNPs in that region will have similar frequencies to each other more often than would be expected under neutrality (see Figure 3.3). In order to compute this summary statistic we have to follow the following steps. First, we choose a genomic window to analyse; then, for every SNP in that window ("core SNP") we calculate the similarity in frequency of the core SNP to every other SNP in the window being analysed. This calculation generates a score that is later exponentiated by a factor *p* (20 in our implementation), which controls the weight given to each SNP, depending on frequency similarity to the "core SNP". This procedure is applied to each SNP in the window and the scores obtained are then averaged, producing a weighted diversity (θ_β) for each window. The Watterson's diversity (Watterson

1975) is then calculated for the window and subtracted to θ_β. The value obtained this way, β, is expected to be approximately 0 in a region under neutrality, and higher than 0 in a region under balancing selection (Siewert and Voight 2017).

For this statistic, in order to minimize the effects of the stochasticity of the occurrence of mutations, we chose a window size of 50Kb that corresponds to the length of the chromosomes we simulated.

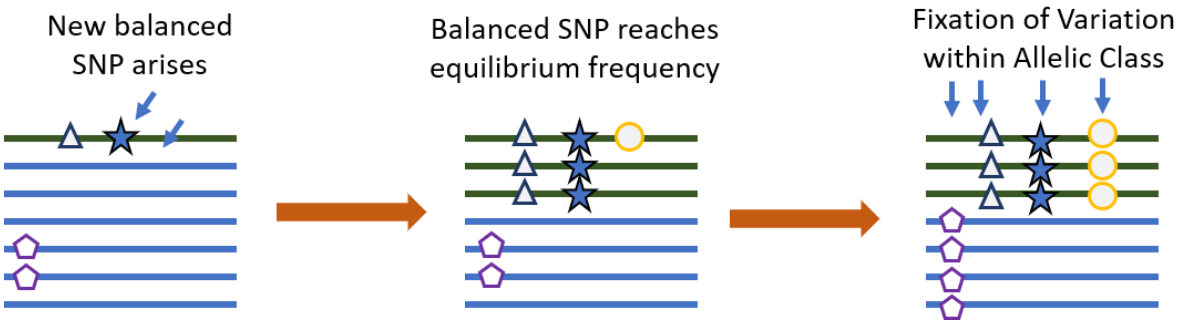

Figure 3.3 ***Schematic of the genomic pattern generated under Balancing Selection.*** *β is based on the correlation between the frequency of an allele under balancing selection (Blue star in the figure) and the frequency of neighbouring alleles. These may co-occur in the same haplotype as the selected allele (the blue triangle and the yellow circle) or on other haplotypes (purple pentagon). Adapted from Siewert and Voight 2017*

4 Results

We start by describing the effects of deleterious mutations in the diversity and differentiation patterns, dividing the results into two sections, first for a single isolated population and then for two populations under isolation with migration models. The results for two populations are divided into two sections that correspond to the scenarios with and without migration between populations. We only show the results for the highest migration rate, as for the other migration rates investigated, we obtained qualitatively similar results to the ones obtained without migration between populations. For simplicity, when summarizing results we use ~ to represent approximately, e.g. ~50% means approximately 50%.

For many of the figures shown in the results, the x-axis corresponds to $\log_{10}(N_e s)$, where N_e is the effective population size and s is selection coefficient of the deleterious mutations (assumed to be equal for all deleterious mutations). For comparison we show the values expected under neutrality, with s=0. However, due to the fact that we plot selective coefficients in log10 scale and $\log_{10}(0)$ is not defined, we arbitrarily defined the s of neutral mutations to be 10^{-7}. As such, the neutral case of s=0 is represented as $\log_{10}(N_e s) = -2$. In order to simplify the text, we refer to selection coefficients between 0 and 0.0001 ($\log_{10}(N_e s)$ between -1 and 1) as "weak selection" because $4N_e$ is much lower than 1, i.e. $s << 1/N_e$; selection coefficients between 0.0001 and 0.01 ($\log_{10}(N_e s)$ between 1 and 2) as "intermediate selection" because they correspond to values of $N_e s$ between 0.1 and 10, i.e. $0.1*(1/N_e) < s < 10*(1/N_e)$; and selection coefficients between 0.01 and 0.1 ($\log_{10}(N_e s)$ between 2 and 3) as "strong selection", as $s \gg \frac{1}{N_e}$.

Given that we want to understand the effects of selection relative to neutral regions, for most figures, the y-axis corresponds to the relative value of each statistic in relation to its neutral expectation, i.e. the ratio between the mean observed value across simulations for each statistic and its neutral theoretical expectation for each combination of parameters. The exceptions to this are the figures that show the absolute results for Tajima's D, R^2 and β, for which values were not standardized relative to the neutral expectation.

4.1 Single Population

In order to verify that SLiM model was implemented correctly, we compared the results of simulations with theoretical expectations of the effect of BGS on linked neutral diversity, which were derived by (Hudson and Kaplan 1995; Campos et al. 2017) for a single population. The results for the comparison between the theoretical expectation for the effects of BGS and the results from SLiM simulations are shown in Figure 4.1 for the two population sizes we tested (N_e=1,000 and 10,000). The theoretical predictions only match the simulations when $\log_{10}(N_e s) > 1$, indicating that simulations do not perfectly match the expected by the theoretical models. This is likely due to the fact that the theoretical predictions models assume no genetic drift and that selection is the main factor. Thus, this is only a good approximation with a sufficiently strong selection, $2N_e s \geq 1$. This hypothesis is corroborated by the results for the bigger population size because in this case the effects of drift are lower and, as such, there is a slightly better agreement between the results from the model and the results from the simulations for low selection coefficients. This indicates that we implemented the BGS model correctly in SLiM.

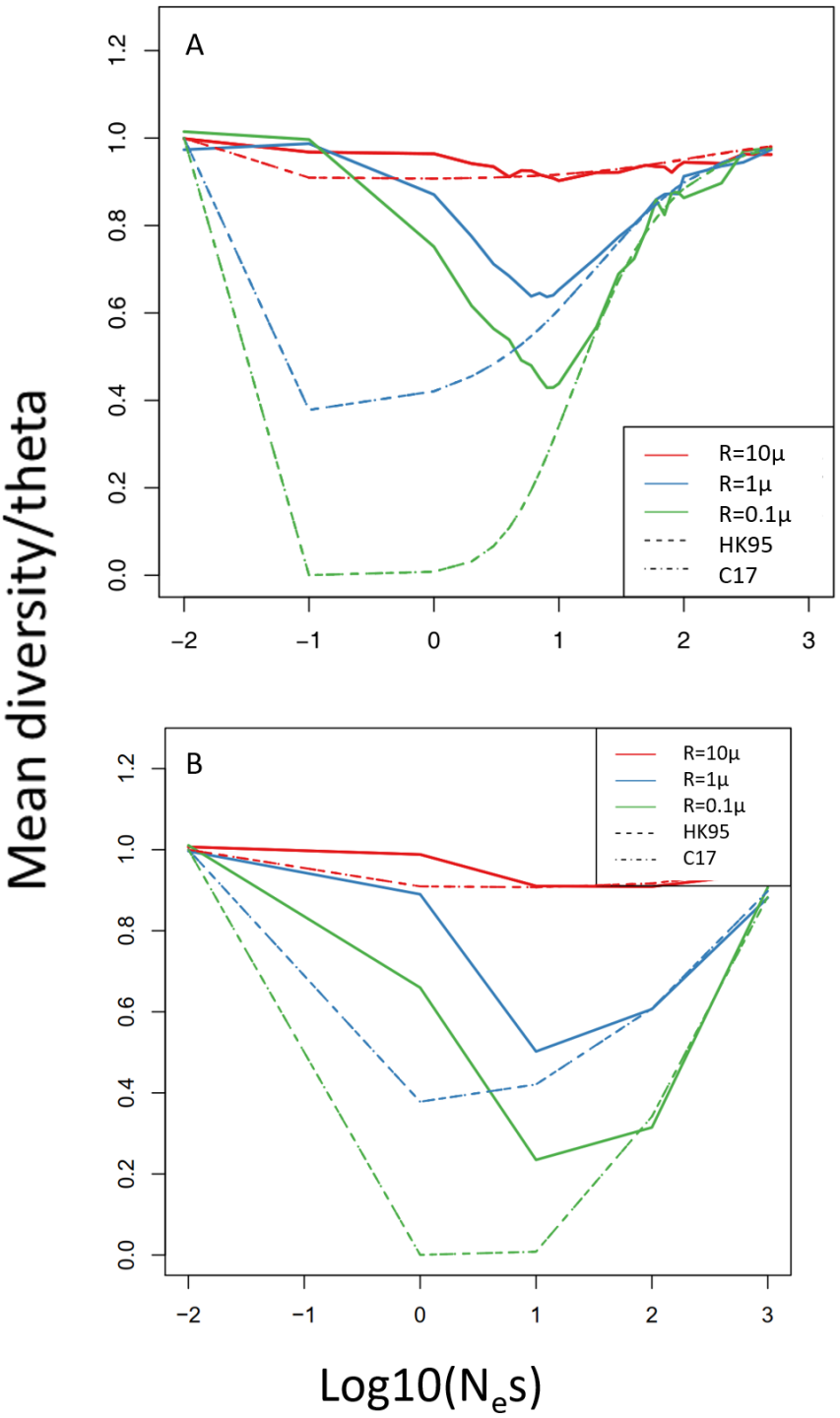

Figure 4.1 ***The effects of linked selection on neutral diversity in a single constant-size population of 1,000 (A) and 10,000 (B) individuals experiencing removal of co-dominant deleterious mutations of fixed selective coefficients, at different recombination rates****. Dotted lines correspond to the theoretical (dotted line) expectation derived for the effects of Background Selection (BGS) by HK95 (Hudson and Kaplan, 1995) and C17 (Campos et al. 2017). Solid lines correspond to the mean of 100 simulations performed using SLiM of a 50Kb region with alternating deleterious and neutral sites with a fixed selective coefficient. Two different lines are shown for the theoretical expectation for the effect of BGS but only one of them is visible because they overlap. The results for a selection coefficient of 0 is ploted at a value of -2, due to the logarithm of 0 being undefined.*

Since there are no theoretical predictions for the neutral value of the β statistic that can be used to detect balancing selection, we computed this statistic for the single population simulations. The results for the statistics to detect balancing selection β (Figure 4.2) show that there is an inverse relationship between the recombination rate and the mean value of this statistic, i.e. the value becomes larger for lower recombination rates. We can also see that, with lower recombination rates there is greater variation in the value of β, likely due to the influence of selection on linked neutral variation. In contrast to the expectation that β=0 for neutral mutations (Siewert and Voight 2017), we found that the value of β depends on the recombination rate. Nevertheless, as expected for neutral mutations, when comparing the results for recessive and co-dominant mutations, their values are very similar for the neutral case with s=0. For the case of the higher and intermediate recombination rates tested, we found that there was no big difference between the β values for co-dominant or recessive mutations. For the case of lower recombination, a difference exists between the values for recessive and co-dominant deleterious mutations for values of selection strength between -0.5 and 2.0 in log10 scale. With recessive mutations we see higher β values than the neutral case for $\log_{10}(N_e s) = -2$, whereas, for co-dominant the β values are lower than neutral case for $\log_{10}(N_e s) = 1$. Values of β higher than the expected under neutrality indicate the action of balancing selection. For strong selection we found β values lower than the neutral case likely indicating no evidence for balancing selection.

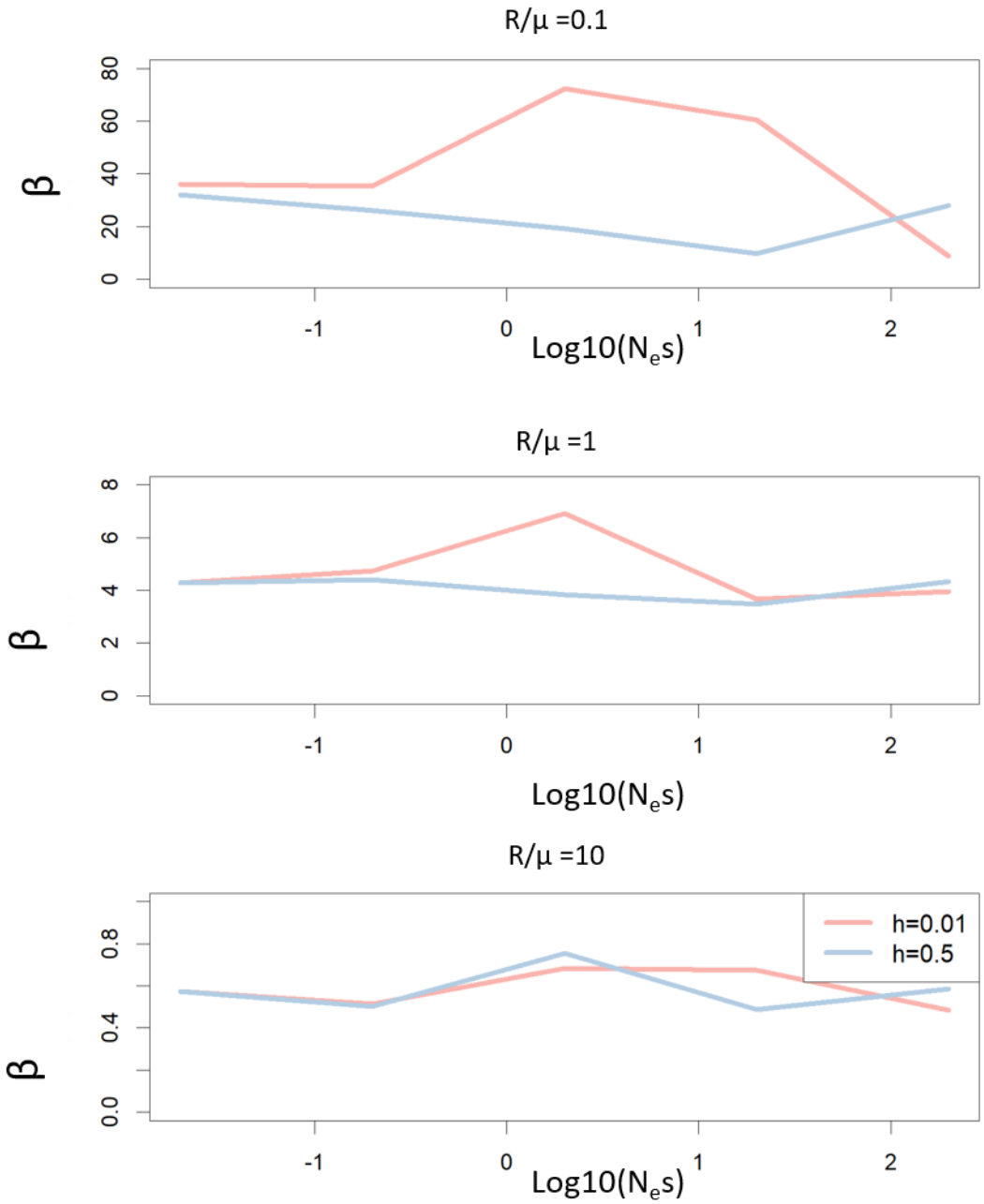

Figure 4.2 ***Effects of linked selection on β in a single constant-size population of 1,000 individuals experiencing removal of co-dominant or recessive deleterious mutations of fixed selective coefficients, with different recombination rates.*** *Solid lines correspond to the mean of 100 simulations performed using SLiM of a 50Kb region with alternating deleterious and neutral sites with a fixed selective coefficient. Note the different scale in the y-axis with each recombination rate. In order to plot the results for neutral mutations (selection coefficient of 0) we show the results for neutrality at -2, due to the logarithm of 0 being undefined.*

4.2 Populations without migration

4.2.1 The effect of co-dominant deleterious mutations

Overall, our results indicate that the removal of co-dominant deleterious mutations affect several statistics that are used to quantify the diversity and differentiation patterns along the genome, and that linked selection can lead to heterogeneous genomic patterns. The effect depends on the selective coefficients and recombination rate.

Regarding the within population diversity (π) of populations under the effects of co-dominant deleterious mutations, the ratio between observed and expected neutral diversity is lower than 1.0 in cases of intermediate selection, indicating that the removal of deleterious mutations decreases diversity of linked neutral variation (Figure 4.3a and b). For weak and strong selection, observed neutral diversity remains close to 1.0, indicating that removal of deleterious mutations has no effect on neutral diversity when mutations are nearly neutral or when selection is strong. The decrease of the ratio between observed diversity and the neutral expectation (θ) that occurs with intermediate selection (between s=0.0001 and s=0.01) is stronger with lower recombination rate (r=0.1μ), with values ranging from approximately 40% to 60% of the neutral expectation, whereas when recombination rate occurs at the same rate as mutation rate (r=μ), the values ranged from 60% to 90% of the neutral expectation. For recombination rates 10 times higher than mutation rate, then the removal of deleterious mutations has a very limited effect on linked neutral diversity, with mean values of relative diversity ranging between ~90% to ~99% of the neutral expectation.

Regarding the absolute genetic distance between populations, for scaled selective coefficients between 0 and 2 in log10 scale, we found that D_{xy} is lower than the neutral expectation, ranging from ~10% to ~20% for the lower recombination and ~10% to ~30% intermediate recombination. This indicates that that the effect of removing deleterious mutations reduces the number of linked neutral mutations, thus reducing the genetic distance between populations (Figure 4.3 c). This is expected because we considered that the effect of removal of deleterious mutations also occurred in the ancestral population, and not only in each population after the divergence. This effect of reduction of D_{xy} is observed for intermediate selection coefficients and regions of lower recombination.

The relative differentiation between the two populations measured by F_{ST} indicates that the removal of deleterious mutations with intermediate selection coefficients leads to increased differentiation, leading to F_{ST} values ~15% to ~30% larger than the neutral expectation (Figure 4.3 d). As seen for diversity, this effect is stronger with lower recombination rate, and F_{ST} values approach the neutral expectation for weak and strong selection. Since F_{ST} measures the relation between the diversity within and between populations, this means that the effect of the reduction in diversity within populations is larger than between populations.

In sum, for models without migration, since the effects of linked selection are stronger with low recombination, we expect regions of the genome experiencing low recombination and affected by co-dominant deleterious mutations to have valleys of diversity and genetic distance and peaks of F_{ST}.

The values of Tajima's D within each population were negative for all selective coefficients, including for neutral mutations with s=0 (Figure 4.3e and f). A negative Tajima's D indicates an increase in the number of rare alleles compared to the expected in neutral sites in a stationary population. As seen with other statistics, the removal of deleterious mutations also affects the value of Tajima's D computed at linked neutral sites, resulting in more negative values indicating an excess of rare alleles. With intermediate selection we found even more negative Tajima's D values, ranging from –0.5 for high recombination rates to -1.2 with low recombination rates. As seen for other statistics, the reduction in Tajima's D values is stronger for the lower recombination rate considered. This pattern, in conjunction with the results for diversity and differentiation, indicates that BGS reduces the diversity of the population by reducing the occurrence of intermediate frequency alleles.

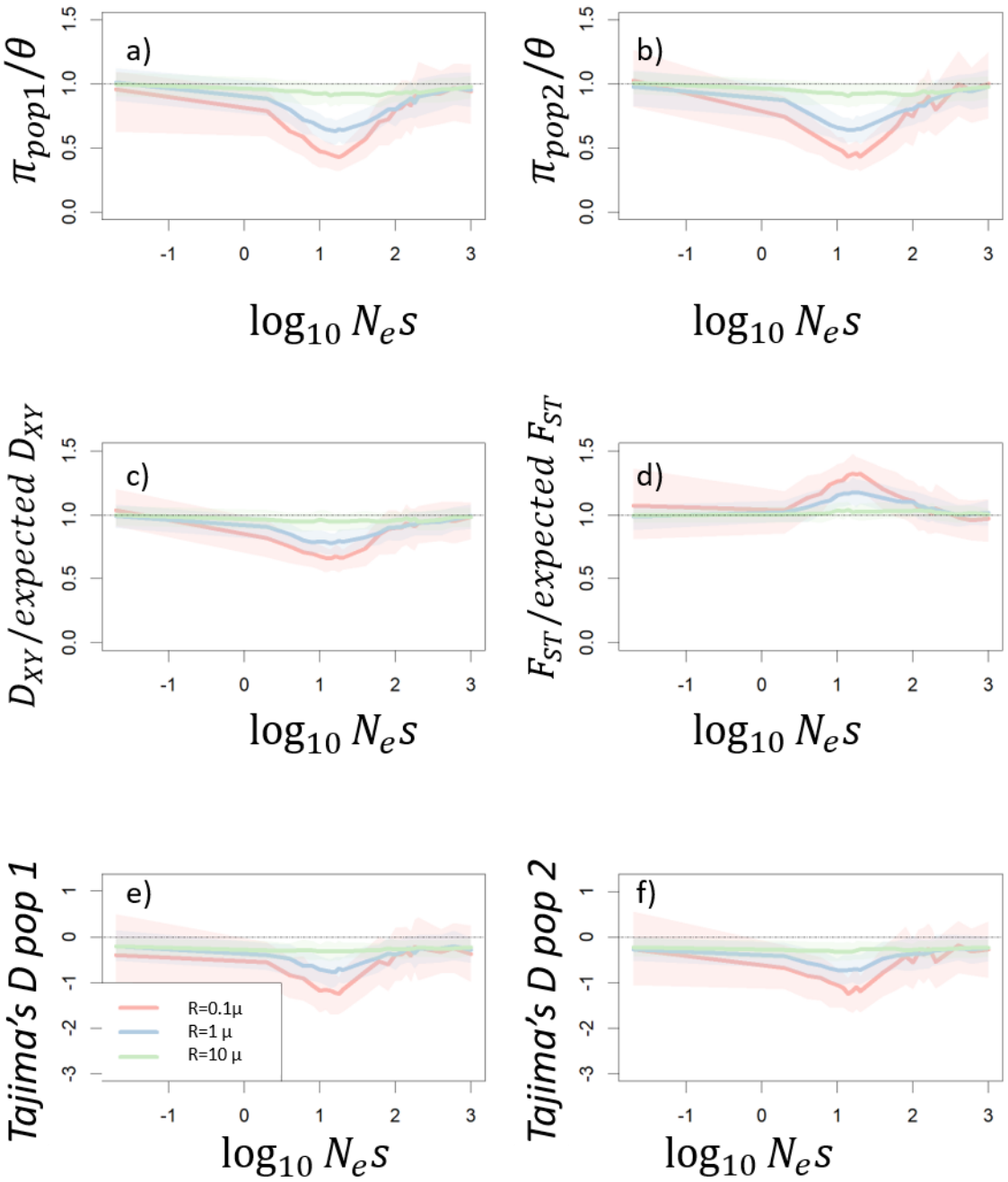

Figure 4.3 ***Measures of various statistics of the neutral mutations for an isolation model with two population and no migration under the effects of co-dominant deleterious mutations****. a) and b) π/(expected π) for population 1 and 2 c) and d)* D_{XY} *and* F_{ST} *between both populations d) and e) Tajima's D for population 1 and 2. In a)-d) the statistics are shown relative to the neutral "expected" value based on analitycal equatiations, with values equal to 1.0 indicating no deviations from neutrality. Results obtained for neutral mutations for 200 simulations performed in SLiM 3.2 for co-dominant deleterious mutation. All these statistics are here represented as a function the logarithm of the selection coefficient for different recombination rates. The solid line represents the mean for the 200 simulations while the shaded area represents the 25% to 75% quantile of the distribution of each statistic across simulations. Different colours correspond to simulations with different recombination rates. Results for neutrality (selection coefficient s=0) are show in log10 scale as -2 as the logarithm of 0 is undefined.*

4.2.2 The effect of recessive deleterious mutations

For the simulations done with recessive deleterious mutations, we find very different patterns when compared with the co-dominant case. The within population diversity (π) was larger than the neutral expectation (θ) for weak and intermediate selection, with values ranging from approximately 50% to 200% for the low and intermediate recombination rates (Figure 4.4 a-b). However, for strong selection (s>0.01, log10(N_es)>2), the diversity was ~50% to ~30% lower than the neutral expectation. The reduction in diversity observed for strong selection is more pronounced for low recombination, which is similar to results of co-dominant mutations.

The genetic distance between populations, measured with D_{xy}, display a similar pattern to the one observed for diversity since it is ~50% to ~200% greater than the neutral expectation with weak and intermediate selection, while also being ~50% to ~30% smaller than neutral expectation with strong selection (Figure 4.4 c). On the other hand, the relative differentiation of populations F_{ST}, shows the reverse pattern, with values of F_{ST} lower than ~50% of the neutral value for intermediate selection, reaching as low as ~10% of the neutral expectation for low recombination rates, and with F_{ST} values ~30% larger than neutral for strong selection (Figure 4.4 d). Thus, for the genetic diversity, genetic distance (D_{xy}) and relative genetic differentiation (F_{ST}) there is a clear transition in the effect of linked selection approximately at $\log_{10}(N_e s)=2$.

For the Tajima's D values, like in the case of co-dominant deleterious mutations, they were also negative for the neutral simulations and with strong selection. However for weak and intermediate selective coefficients we found positive Tajima's D, reaching values between 0.5 and 1.0 for the simulations with lower recombination rate (Figure 4.4 e and f). Together with results for diversity, this indicates that the increase in neutral diversity seen with weak and intermediate selection is not accompanied by an increase in the number of segregating sites with rare allele frequencies. Instead, positive Tajima's D indicates a deficit of rare mutations, and an increase in intermediate allele frequencies, which may occur due to balancing selection or due to demographic events, such as population contraction and population structure.

4.2.3 The Effect of linked selection on Linkage Disequilibrium (LD)

Although our simulations with different selective coefficients assume a constant recombination rate across each chromosome of 50Kb, the effects of linked selection can affect the patterns of linkage disequilibrium (LD), which we measured with the r^2 statistic. Values of r^2 close to 1.0 indicate strong LD and values close to 0 indicate no LD. We show the average r^2 for each population in Figure 4.5. Since lower recombination rates lead to a lower probability of crossing-over occurring between contiguous sites, a higher value of linkage disequilibrium (and of r^2) is expected with lower recombination rates. Thus, the values of r^2 depend directly on the recombination rate, even under neutrality, and hence we compare the effects of selection for each recombination rate separately.

We found that the r^2 with deleterious co-dominant mutations, with low recombination rate, decreases by ~40% below the neutral value, reaching the minimum with $\log_{10}(N_e s) = 1$. For stronger selection r^2 increases, but it is still ~20% below neutral values. For intermediate recombination, r^2 decreases up to ~5% below the neutral value with $\log_{10}(N_e s) = 1$ but increases up to ~10% above neutral value with strong selection of $\log_{10}(N_e s) = 3$. In both scenarios the turning point where r^2 begins to increase occurs when selection becomes stronger than drift, at $\log_{10}(N_e s) > 1$. One possible explanation is that for selection coefficients lower than the turning point, deleterious mutations may increase in frequency

due to drift. This in turn will lead the co-occurrence of deleterious mutations in the same individual but in a heterozygote state and in different haplotypes to become more common. Thus, recombination will produce haplotypes with less deleterious mutations and these are favoured by selection. This situation is reversed if selection is stronger than drift because in that case selection efficiently removes deleterious mutations and linked neutral variation. This decreases diversity and increases LD.

When it comes to recessive deleterious mutations the r^2 is greater than the neutral value with low selection and intermediate or low recombination, increasing by ~30% and ~5% respectively. Thus, we can conclude that, in these scenarios, non-recombinants are favoured by AOD. For values of *s* larger than $\log_{10}(N_e s) > 2$ we can see a decrease in the value of r^2 bellow neutral values. This is the same as found in the case of co-dominant mutations, indicating a transition from AOD to BGS. In sum, AOD favours non-recombinants (high LD) while BGS favours recombinants (low LD).

High recombination rate leads to low values of r^2 and in that case we found similar results for recessive and co-dominant mutations. In this scenario we can see that varying the selection coefficient against deleterious mutations does not have a strong effect on the values of r^2, indicating that with high recombination selection does not affect LD patterns.

4.2.4 The effect of linked selection on statistics to detect Balancing Selection (β)

We show the values of the statistic β used to detect balancing selection for each population and recombination rate in Figure 4.6 The results for the high recombination show that, if recombination is frequent enough, β is independent of selection. However, with intermediate and low recombination rates there is a high degree of variation in the β score due to changes in the selection coefficient. In the case of co-dominant mutations, there is a reduction of β with intermediate selection below neutral levels, a sign that balancing selection is not occurring. However, with recessive mutations, the values of β are 100% higher than neutral expectation with intermediate selection. Similar to the results in a single population, we found that the value of the β statistic is never 0 for the neutral case, as expected, and is instead dependent on the recombination rate.

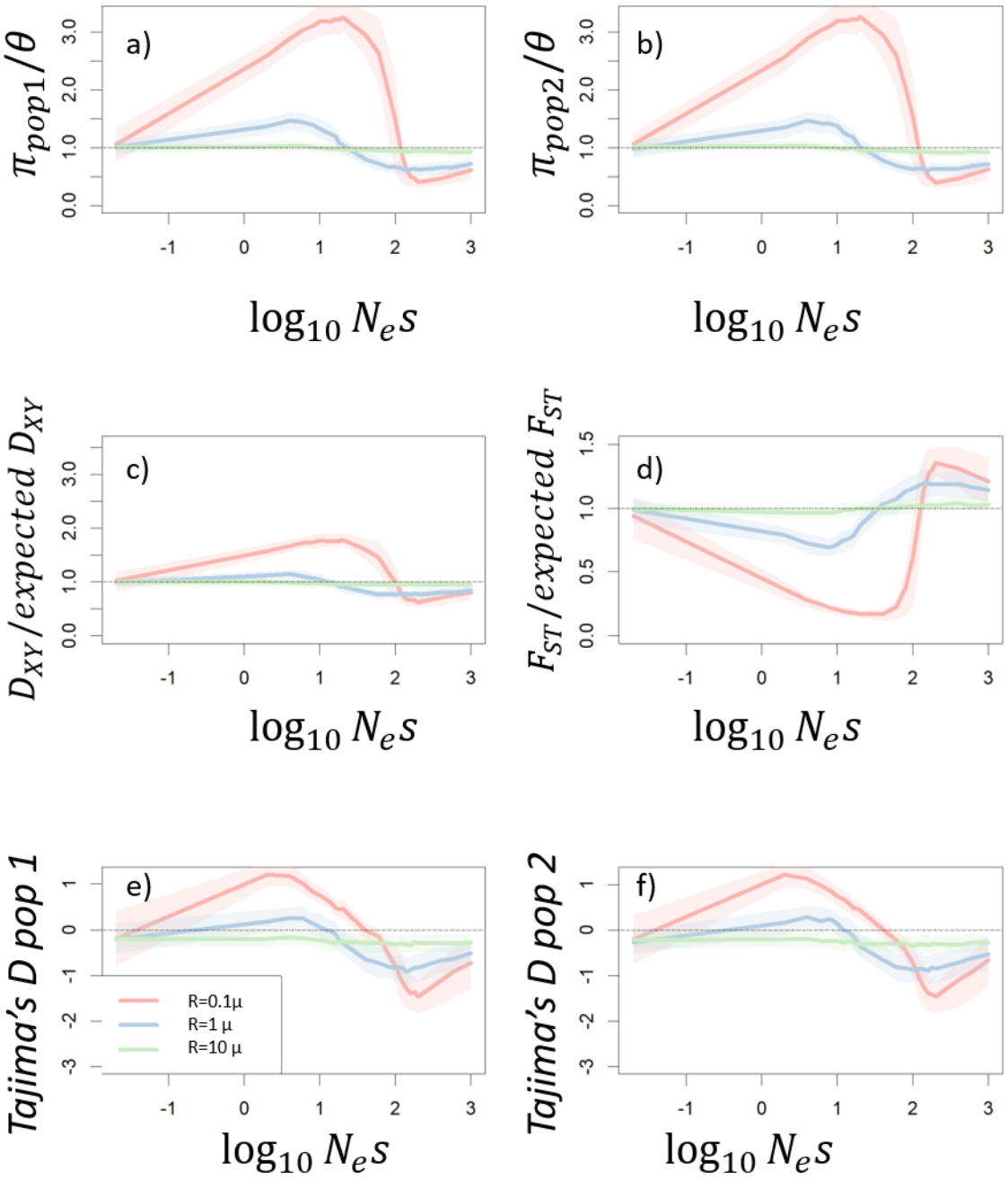

Figure 4.4 ***Measures of various statistics for an isolation model with two population and no migration under the effects of recessive deleterious mutations*** *a) and b) π/(expected π) for population 1 and 2 c) and d) D_{xy} and F_{ST} between both populations d) and e) Tajima's D for population 1 and 2 for neutral mutations. In a)-d) the statistics are shown relative to the neutral "expected" value based on analitycal equatiations, with values equal to 1.0 indicating no deviations from neutrality. Results obtained by performing 200 simulations in SLiM 3.2 for each parameter combination. All these parameters are here represented as a function of the logarithm of the selection coefficient for different recombination rates. The solid line represents the mean for the 200 simulations while the shaded area represents the 25% to 75% quantile of the distribution of each statistic across simulations. Different colours correspond to simulations with different recombination rates. Results for neutrality (selection coefficient s=0) are show in log10 scale as -2 as the logarithm of 0 is undefined.*

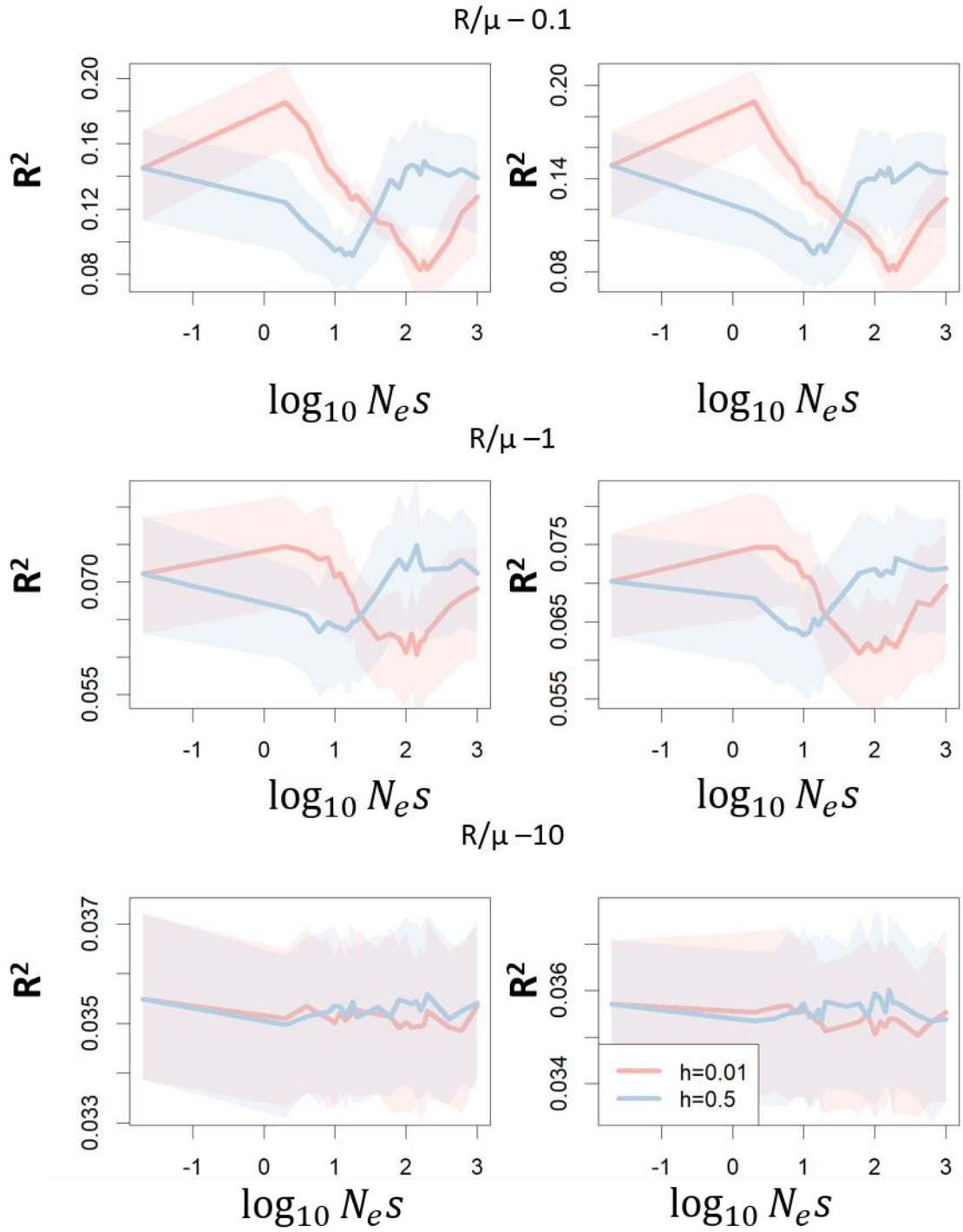

Figure 4.5 ***The effect of linked selection on linkage disequilibrium (r^2) in an isolation model with two constant-size populations of 1,000 individuals and no migration, experiencing removal of co-dominant or recessive deleterious mutations of fixed selective coefficients, with different recombination rates.*** *Solid lines correspond to the mean of 200 simulations performed using SLiM of a 50Kb region with alternating deleterious and neutral sites with a fixed selective coefficient. Note the different scale in the y-axis with each recombination rate. The solid line represents the mean for the 200 simulations while the shaded area represents the 25% to 75% quantile of the distribution of each statistic across simulations. Different colours correspond to simulations with different dominance coefficients h. Results for neutrality (selection coefficient s=0) are show in log10 scale as -2 as the logarithm of 0 is undefined.*

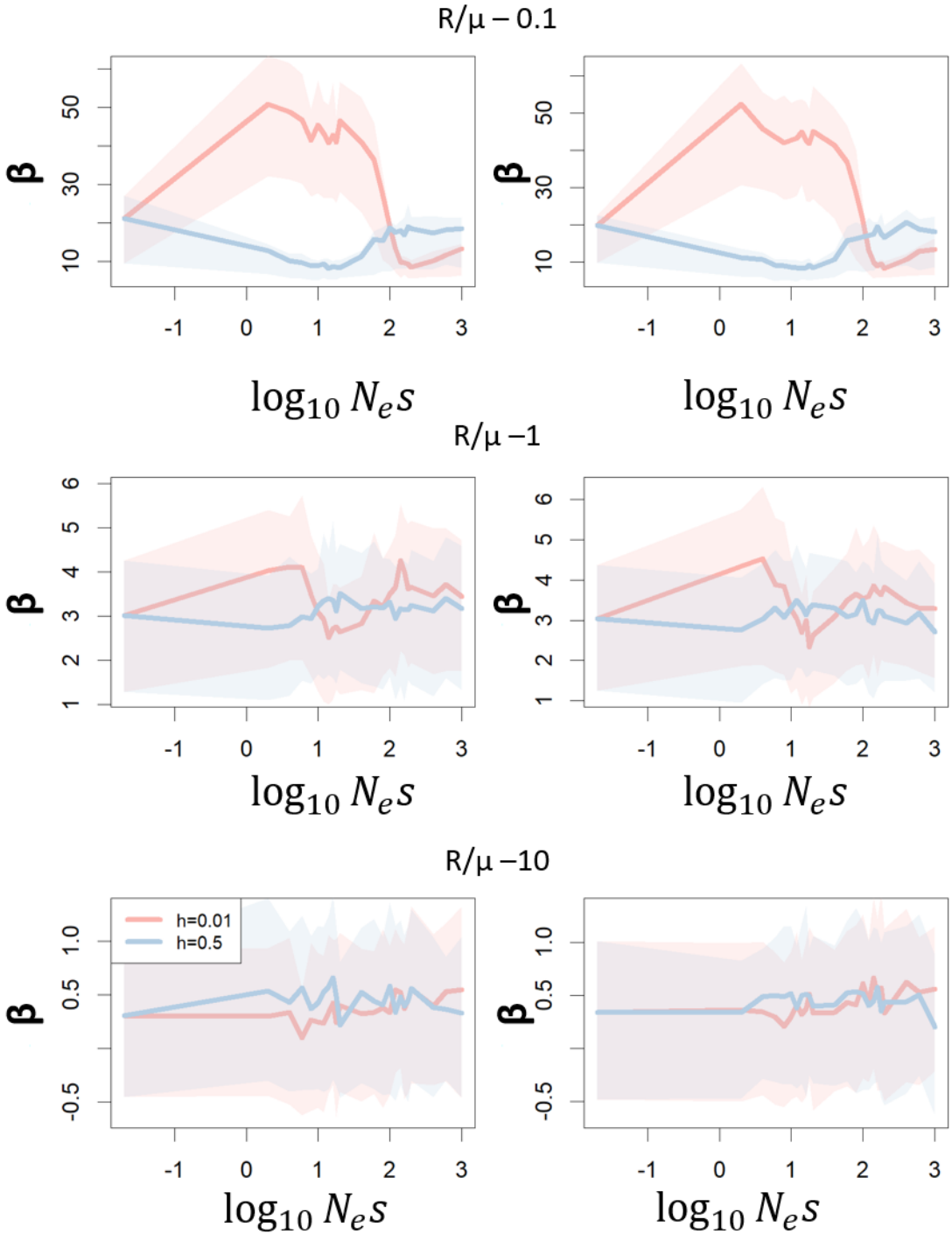

Figure 4.6 ***The effect of linked selection on statistics to detect balancing selection (β) in two constant-size population of 1,000 individuals experiencing removal of co-dominant or recessive deleterious mutations of fixed selective coefficients, with different recombination rates****. Solid lines correspond to the mean of 200 simulations performed using SLiM of a 50Kb region with alternating deleterious and neutral sites with a fixed selective coefficient. Note the different scale in the y-axis for each recombination rate. The solid line represents the mean for the 200 simulations while the shaded area represents the 25% to 75% quantile of the distribution of each statistic across simulations. Different colours correspond to simulations with different dominance coefficients h. Results for neutrality (selection coefficient s=0) are show in log10 scale as -2 as the logarithm of 0 is undefined.*

4.3 Populations experiencing migration

Here, we show only the results concerning the isolation with migration models under the highest migration rate ($2N_em = 10$), and we show results for the other migration rates considered ($2N_em = 1$ and 0.1) in the Supplementary Figures S1 through S8. The reason for this separation is that results for the two lower migration rates do not differ qualitatively from the results without migration between the two populations.

4.3.1 Effect of migration with deleterious co-dominant mutations

We show the results for co-dominant deleterious mutations in Figure 4.7. A seen in the figure, the effects that deleterious mutations have on diversity or the Tajima's D are the same with migration or without migration (as seen on Figure 4.1) but we found differences in results for D_{xy} and F_{ST}. The D_{xy} shows a reduction below neutral expectation (between ~5% with high recombination to ~50% with low recombination). The F_{ST} on the other hand, shows a noisy pattern with most of the relative values falling around the neutral expectation of 1.0. Taken together these results show that strong migration ($2N_em = 10$) reduces the genetic distance between the two populations and erases the increase in F_{ST} that is generated without migration and with low migration rates withco-dominant mutations (Figure 4.3).

4.3.2 Effect of migration with deleterious recessive mutations

We show the results for diversity with recessive mutations in Figure 4.8. As can be seen in the figure, we found the same pattern as the one that occurs in populations without migration: an increase of ~150% in diversity when compared with neutral theoretical expectation with intermediate selection and a decrease of ~50% with strong selection. However these effects are not as pronounced in the case of high migration. The results for D_{xy} do not differ qualitatively from the ones observed without migration. However, there is an increase of ~140% in the observed value for intermediate selection values, bigger than the increase observed with same recombination rate and selective coefficient without migration. This suggests that linked selection leads to an increase in observed D_{xy} that is stronger than the reduction of the neutral theoretical expectation of the D_{xy} that is caused by an increase in the migration rate. The F_{ST} shows the same pattern as seen in the case of populations without migration, ~60% smaller than the neutral expectation with intermediate recombination and intermediate selection, being approximately equal to the neutral expectation with weak and strong selection. This indicates that the effect that deleterious linked selection has on differentiation is independent of the occurrence of migration between populations. However, this pattern shows some noise caused by the stochasticity of the simulation process.

4.3.3 Effect of migration and linked selection on linkage disequilibrium (LD) patterns

Our results show that the linkage disequilibrium, as measured with r^2, is slightly reduced by migration when compared with the neutral result (Figure 4.9). This effect is seen with all recombination rates and dominance coefficients. However, when compared with the results for populations diverging without migration, the results do not show a marked qualitative difference, indicating that migration does not alter the main effect of linked selection. This means that the effects that linked selection has on LD are not greatly changed by the occurrence of migration.

4.3.4 Effect of migration and linked selection on statistics to detect balancing selection (β)

When compared to the case of isolated populations without migration, we found that the β statistic values are reduced by migration, although this effects is only seen in the case of low and high recombination

(Figure 4.10). In the case of high recombination there is a considerable decrease of the maximum value of this statistic (~50%), while with low recombination this reduction is less pronounced (~10%). In the case of intermediate selection there is an increase in the maximum value due to the occurrence of migration of about 10%. Despite these changes in magnitude, the results with migration are qualitatively similar to the ones obtained in the absence of migration, suggesting that the occurrence of balancing selection with recessive mutations is independent of migration.

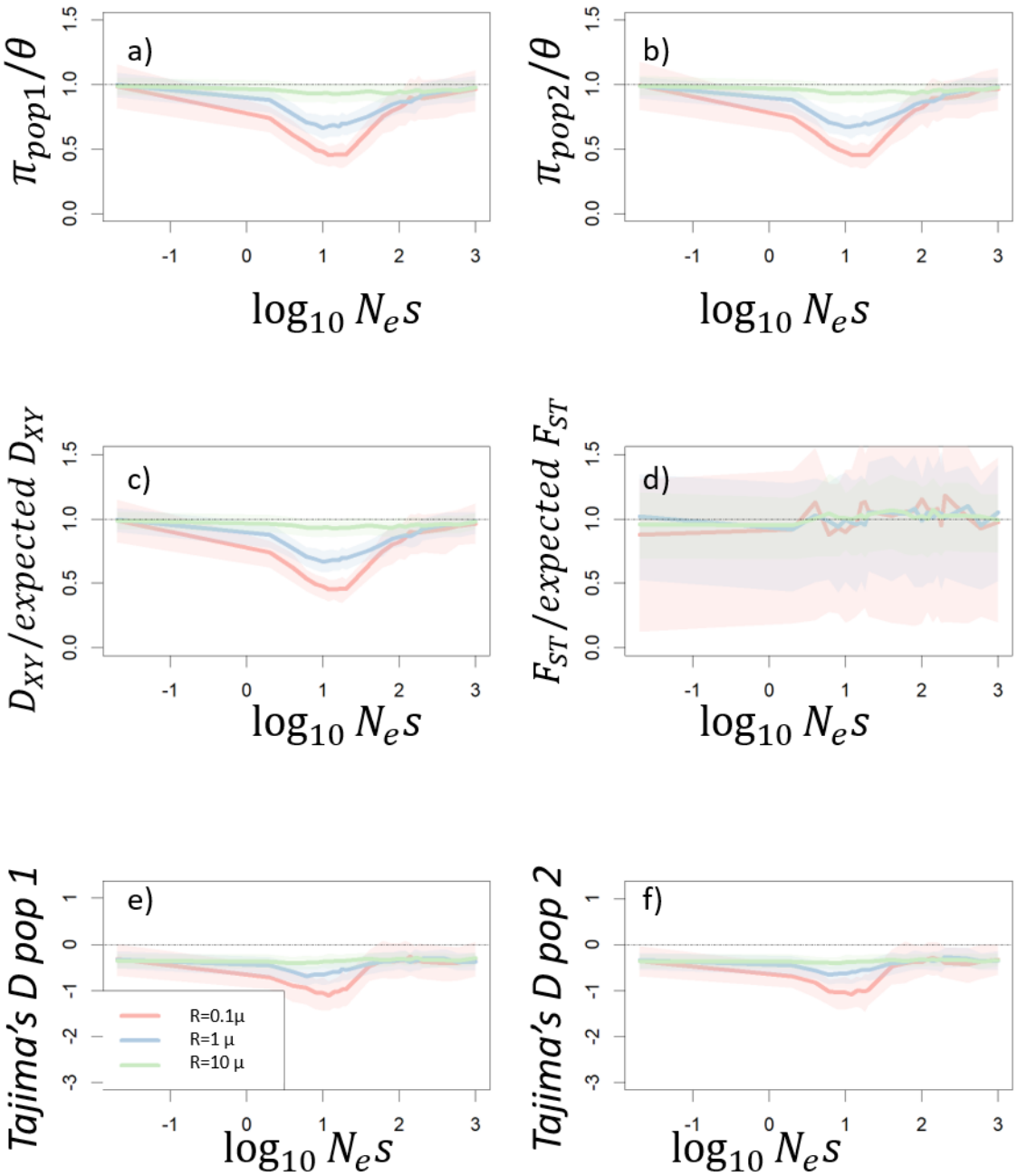

Figure 4.7 ***Measures of various statistics for an isolation with migration model with two population and constant migration ($2N_em=10$) under the effects of co-dominant deleterious mutations*** *a) and b) π/(expected π) for population 1 and 2 c) and d) D_{xy} and F_{ST} between both populations d) and e) Tajima's D for population 1 and 2 for neutral mutations. In a)-d) the statistics are shown relative to the neutral "expected" value based on analitycal equatiations, with values equal to 1.0 indicating no deviations from neutrality. Results obtained by performing 200 simulations in SLiM 3.2 for each parameter combination . The solid line represents the mean for the 200 simulations while the shaded area represents the 25% to 75% quantile of the distribution of each statistic across simulations. Different colours correspond to simulations with different recombination rates. Results for neutrality (selection coefficient s=0) are show in log10 scale as -2 as the logarithm of 0 is undefined.*

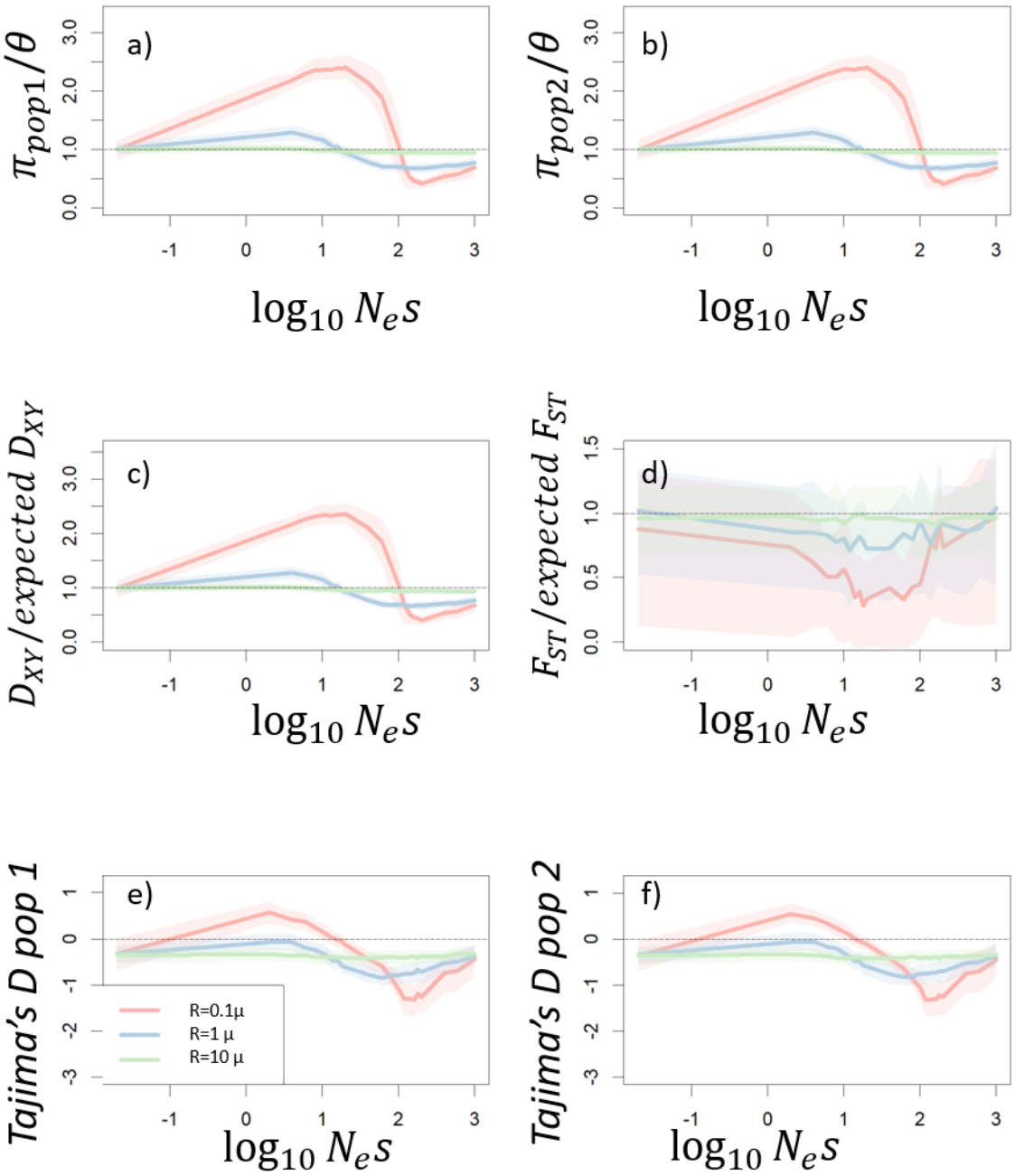

Figure 4.8 ***Measures of various statistics for an isolation with migration model with two populations and constant migration ($2N_em=10$) under the effects of recessive deleterious mutations.*** *a) and b) π/(expected π) for population 1 and 2 c) and d) D_{xy} and F_{ST} between both populations, respectively d) and e) Tajima's D for population 1 and 2 for neutral mutations. In a)-d) the statistics are shown relative to the neutral "expected" value based on analitycal equatiations, with values equal to 1.0 indicating no deviations from neutrality. Results obtained by performing 200 simulations in SLiM 3.2 for each parameter combination. The solid lines represents the mean for the 200 simulations while the shaded area represents the 25% to 75% quantile of the distribution of each statistic across simulations. Different colours correspond to simulations with different recombination rates. Results for neutrality (selection coefficient s=0) are show in log10 scale as -2 as the logarithm of 0 is undefined.*

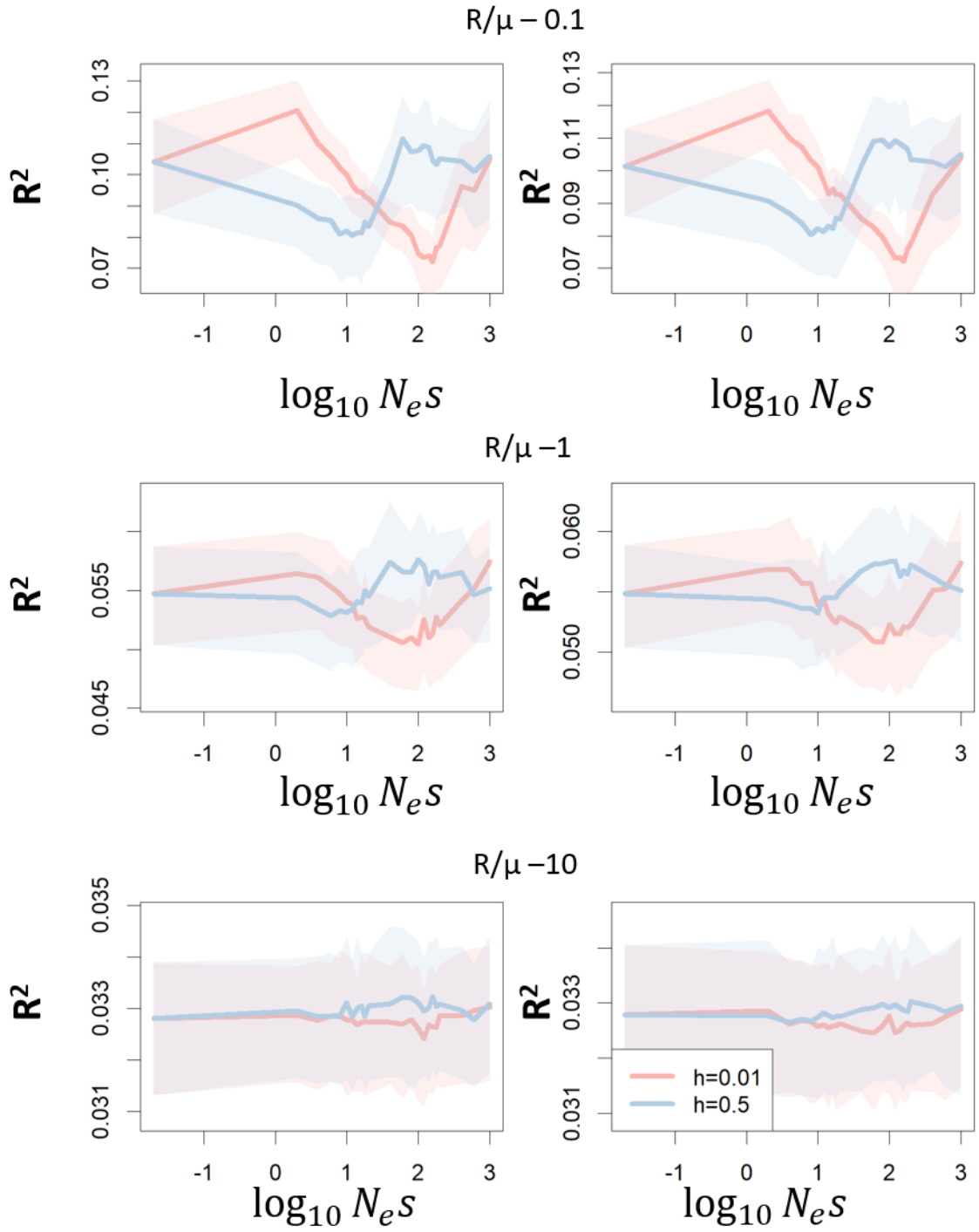

Figure 4.9 ***The effect of linked selection on linkage disequilibrium (r^2) in an isolation with migration model with two constant-size populations of 1,000 individuals and constant migration ($2N_em=10$), experiencing removal of co-dominant or recessive deleterious mutations of fixed selective coefficients, with different recombination rates.*** *Solid lines correspond to the mean of 200 simulations performed using SLiM of a 50Kb region with alternating deleterious and neutral sites with a fixed selective coefficient. Note the different scale in the y-axis with each recombination rate. The solid line represents the mean for the 200 simulations while the shaded area represents the 25% to 75% quantile of the distribution of each statistic across simulations. Different colours correspond to simulations with different dominance coefficients h. Results for neutrality (selection coefficient s=0) are show in log10 scale as -2 as the logarithm of 0 is undefined.*

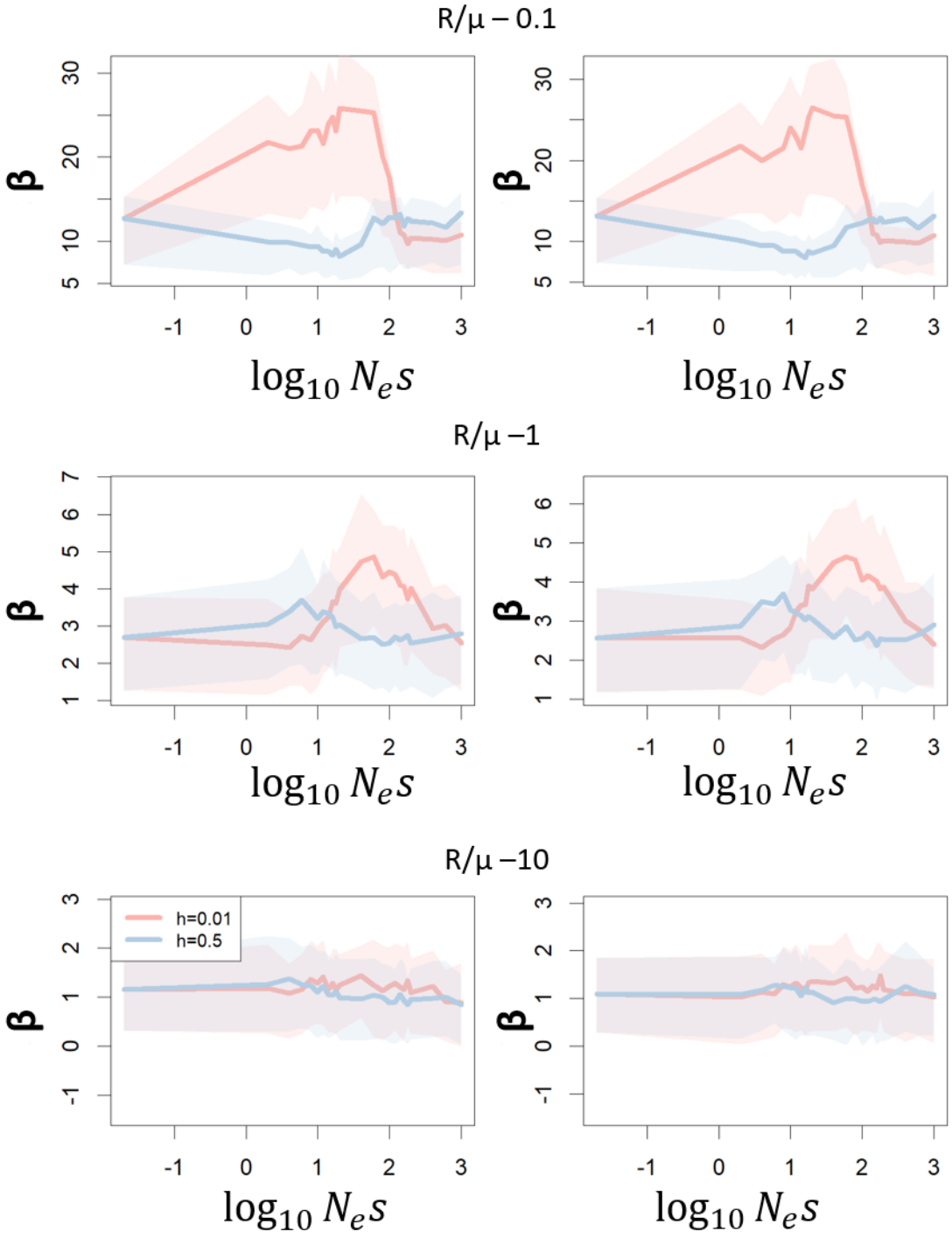

Figure 4.10 ***The effect of linked selection on a statistic to detect balancing selection (β) in an isolation with migration model with two constant-size population of 1,000 individuals and constant migration ($2N_em=10$) experiencing removal of co-dominant or recessive deleterious mutations of fixed selective coefficients, with different recombination rates***. *Solid lines correspond to the mean of 200 simulations performed using SLiM of a 50Kb region with alternating deleterious and neutral sites with a fixed selective coefficient. Note the different scale in the y-axis for each recombination rate. The solid line represents the mean for the 200 simulations while the shaded area represents the 25% to 75% quantile of the distribution of each statistic across simulations. Different colours correspond to simulations with different dominance coefficients h. Results for neutrality (selection coefficient s=0) are show in log10 scale as -2 as the logarithm of 0 is undefined.*

5 Discussion

5.1 Co-dominant and recessive deleterious mutations lead to different genomic patterns

We studied the interaction between linked purifying selection and migration in the context of the isolation with migration model, simulating data with an individual-based forward-in-time program, SLiM (Haller and Messer 2018). We observed that co-dominant mutations lead to a decrease of within-population diversity with intermediate selection coefficients. However, with recessive mutations there is a transition between AOD to BGS.

In the case of BGS, in conditions where its effects were maximized, we found a decrease in diversity below the neutral value of approximately 45 % in models of a single isolated population. For single population models, here are theoretical results for the effects of BGS on linked neutral diversity (Hudson and Kaplan 1995; Campos et al. 2017). We compared those theoretical expectations with the results obtained from our simulations and found that, when the selective coefficients *s* is lower than 0.001, there are deviations. One of the possible explanations for this is that the models assume that deleterious mutations never rise in frequency above their equilibrium in the absence of genetic drift. However, with low selection coefficients, the mutation-selection-drift equilibrium changes and drift becomes stronger than selection, leading to a possible increase in frequency of deleterious alleles. This explanation is consistent with our results, as relative diversity deviated from the theoretical expectation when $N_e s<10$. As expected, for co-dominant mutations with $N_e s>1000$ the linked diversity is not affected by selection, with values similar to the neutral expectation, which is explained by the fact that when deleterious mutations have a strong effect they are immediately eliminated from the single haplotype they occur on, without increasing in frequency, and hence without affecting neutral linked diversity.

The decrease in diversity that we found is consistent with the effects of BGS. This is also seen in the case of an isolation model with two populations and no migration between them (~50%). We also found that BGS leads to a reduction of the D_{xy} of up to ~30%, while increasing the F_{ST} up to ~30%. The occurrence of BGS is also associated with a reduction of Tajima's D to values close to -1.0 and a reduction in the r^2 of approximately 50% compared to the neutral expectation. Previous studies, similar to ours, also found that BGS leads to a deviation from the expected neutral SFS, affecting the Tajima's D values (Cvijović et al. 2018; Matthey-Doret and Whitlock 2019).

In contrast, for slightly recessive deleterious mutations we found an increase in diversity between 100 and 200%, in genetic distance between ~50 and ~100%, Tajima's D between 0 and 1.1 and r^2 of approximately 30% as well as a reduction in F_{ST} of approximately 70%, which is consistent with the occurrence of AOD. The increase in diversity caused by AOD has been reported in other theoretical studies (Pamilo and Pálsson 1998; Paælsson 2004; Gilbert et al. 2019). Gilbert et al. 2019 performed a simulation study and developed a model of AOD for a single population, and found similar effects on diversity, leading to an increase of up to 200% of the neutral expected value, which matches very closely what we found. Gilbert et al. 2019 results indicate that AOD only occurs if the ratio between the recombination rate and the mutation rate is lower than 0.34. However, we found an increase in diversity and reduction in differentiation that is consistent with the effects of AOD occurring even when the recombination to mutation rate ratio is 1.0. These differences can be because we considered 50% of mutations to be deleterious, whereas Gilbert et al. 2019 considered 2% of deleterious mutations.

A reduction in diversity and D_{xy} together with an increase in F_{ST} due to BGS were described very recently in a simulation study by (Matthey-Doret and Whitlock 2019). These authors found a significant

but smaller decrease of diversity (~5 % of neutral value) than we found in the same combinations of parameters. These authors performed a simulation study of the effect of co-dominant deleterious mutations under various migration rates. In the absence of migration or with low migration rates ($2N_em$=1) they found that BGS leads to a significant increase of F_{ST} of approximately 5% in both scenarios, while higher migration rates (equal to our maximum of $2N_em$=10) lead to the a non-significant reduction of F_{ST}. This is similar to what we found, which was an increase of F_{ST} of ~30% and no effect for higher migration rate of $2N_em$=10. Overall, we found similar results, but there were some differences. Because one of the main factors that influences the divergence is the relationship between population size and divergence time, and given that the migration rate and the population sizes are the same in both studies, we can conclude that the differences between both studies are related to differences in the divergence times and dominance coefficients. In fact, in our simulations we considered a split time of 2000 generations (t_0=2N_e), while (Matthey-Doret and Whitlock 2019) set this parameter to 5000 (t_0=5N_e) generations, leading to a greater divergence in that study.

Regarding the measurements of balancing selection our results for β show that there is a reduction of this statistic when BGS occurs but that there is an increase of β when AOD occurs. This suggest that the genomic patterns generated under AOD are similar to the ones generated under balancing selection, and that this statistic can be useful to detect regions under AOD. However, this results should be interpreted with caution, due to the fact that we found that the value of β depends on the recombination rate and it was different from the expected neutral value of zero (Siewert and Voight 2017) in our neutral simulations. Our results also show that Tajima's D is negative in the neutral case, suggesting that there is a systematic bias towards slightly negative values of Tajima's D. This is an unexpected result because, in the absence of population bottlenecks, expansions or selection, Tajima's D is expected to be 0. We need to further investigate if this is due to an error in our script to compute the Tajima's D, by testing it with a different simulator of neutral diversity, such as coalescent simulations.

Consistent with previous studies, we found that the extent and magnitude of the effects that linked selection has on the genome are dependent on two key factors: the distribution of fitness effects and the distribution of dominance effects of deleterious mutations. Mutation accumulation studies from *Drosophila melanogaster* (García-Dorado et al. 1998) and *Saccharomyces cerevisiae* (Agrawal and Whitlock 2011) suggest that deleterious mutations are usually recessive and the mean selective coefficient of deleterious mutation in these organisms lies between 0.1 and 0.3 for *Drosophila* and approximately 0.045 for *S. cerevisiae*. This suggests that that the conditions conductive to the occurrence of AOD ($N_e s \leq 100$ according to our results) will be common as long as the effective population sizes are smaller, less than 1000 for *D. melanogaster* and less than 10,000 for *S. cerevisiae* or if selective coefficients are low.

Previous studies of gene deletions in *S. cerevisiae* also found that there is an inverse relationship between the dominance coefficient and the selection coefficient, with lower dominance for mutations with a stronger effect (Phadnis and Fry 2005). This suggests that mutations of small effect will tend to have higher dominance coefficients, and hence could imply that conditions that generate AOD (recessive mutation of small effect) are rare. Still, the relationship between selective coefficient and dominance was dependent on the environment and most deleterious mutations were estimated to be partly recessive (Phadnis and Fry 2005). There are several lines of evidence for the importance of slightly deleterious and recessive mutations. First, studies where the distribution of fitness effects is inferred suggest that most deleterious mutations that are not lethal have a slight fitness effect (Eyre-Walker and Keightley 2007). Second, studies from mutation accumulation and knockout mutations in model organisms suggest that mildly deleterious mutations are partially recessive (Agrawal and Whitlock 2011). Third, recent

estimates using genetic polymorphisms and comparative genomics indicate that many genes are affected by deleterious mutations. For instance, (Huang and Siepel 2019) found that in humans, approximately 2000 genes show evidence of negative selection, with 773 under weak negative selection. Another example comes from genomic data from modern and ancient DNA from domestic species, suggesting that mutation load in domestic horses is mostly due to slightly deleterious and recessive mutations (Orlando and Librado 2019). Fourth, very recent studies found evidence for AOD in genomic regions of low recombination in humans (Gilbert et al. 2019) and *Drosophila* (Schou et al. 2017).

5.2 The Effect of BGS and AOD are maximized for low recombination rates

Our results show that for most selective coefficients and parameter combinations, the statistics were similar to their corresponding neutral expectation with the higher recombination rate tested. This is an expected result because both in the case of BGS (Hudson and Kaplan 1995) and in the case of AOD (Pamilo and Pálsson 1998) the strength of the effect is inversely related to the recombination rate and positively related to the mutation rate. Note that the mutation rate was fixed in our simulations, but since we set the recombination rate relative to the mutation rate, our results are general.

For BGS, the lower the recombination rate, the lower the diversity and genetic distance and the higher the F_{ST}. The opposite is true for AOD, that is, the lower the recombination, the higher the diversity and genetic distance and the lower the $F_{ST.}$ We also found that the occurrence of AOD leads to an increase in LD, as seen by the results for the r^2. On the other hand, the occurrence of BGS leads to a decrease of BGS. The explanation for the pattern we found lies in the fact that, BGS leads to recombinant haplotypes having greater fitness than non-recombinants and, therefore, the r^2 we found was lower than neutral. The opposite was found in the case of AOD, that is, it leads to a greater fitness for non-recombinant haplotypes. This is something unexpected that should be explored in further studies, such as scenarios without recombination and scenarios with epistatic interactions.

Because the recombination rate is the main determinant of the magnitude of the effects of linked selection, and there is variation of the recombination rate along the genome (Comeron et al. 2012), this leads to the occurrence of variation of the magnitude of BGS and AOD along the genome as predicted previously for BGS (Nordborg et al. 1996). This also means that, due to their "neutral-like" behaviour, one can use zones of the genome with high recombination rate as a control for the occurrence of AOD and BGS.

5.3 The Effects of BGS and AOD are mostly independent of migration rate

Our results show that migration does not fundamentally change the impact that linked selection has on the relative genomic patterns generated in models without migration. Our results indicate that with co-dominant deleterious mutations there is an increase in F_{ST} of up to ~30% in regions of low recombination with migration rates below $2N_em$ <1.0, consistent with the effects of BGS. This is very similar to the strongest effect that is seen without migration, where also we found an increase in F_{ST} of ~30%. This has been predicted by some theoretical models that have suggested that BGS may increase F_{ST} in populations with (Zeng and Corcoran 2015) and without migration (Charlesworth et al. 1997; Nordborg 1997).

This effect can be understood in terms of the effects of BGS on the effective size. (Hudson and Kaplan 1995) showed that BGS can be approximated as a local reduction in the effective size (N_e). As a result of a lower N_e than neutral genomic regions, regions experiencing BGS are expected to have a lower

genetic diversity and higher genetic differentiation. However, we found that for migration rates of $2N_em$=10.0 the values of F_{ST} do not seem to be affected by BGS, with noisy values around 1.0 for most selective coefficients. This suggests that when migration is strong enough the increase in F_{ST} due to BGS is eliminated. This can be explained by the fact that migration reduces the within (π) and between (D_{xy}) population pairwise distance in the same proportion. However, we note that the results are noisy, especially because with a $2N_em$=10.0 we expect already a very low F_{ST} of 0.02. Thus, we would need to further investigate if this result holds for lower migration rates or for higher divergence times.

In contrast, for recessive slightly deleterious mutations we found an increased genetic diversity of at least ~100% and reduction in F_{ST} of at least ~80% for N_es<100 for all migration rates tested, including the case of no migration. Thus, the patterns of genomic signatures generated by AOD, in contrast to the case of BGS, is not erased by strong migration. This suggests that if AOD is the main process, we should find its genomic signatures in regions of low recombination in populations with or without gene flow. In contrast, for regions of the genome where BGS is the main process, we expect to find differences between populations experiencing low and high levels of gene flow. This prediction could be tested with data from natural populations from related species experiencing different levels of gene flow.

5.4 The Transition from AOD to BGS with recessive mutations

Dependent on the selective coefficient of recessive deleterious mutations, our results indicate that there is a transition between BGS and AOD. Interestingly, this transition was consistently seen at values of N_es~100 even when varying levels of gene flow and for the low and intermediate recombination rates considered.

This transition is observed in all the statistics we measured: diversity π, D_{XY}, F_{ST}, Tajima's D, r^2 and β. For diversity we see a transition from at least 200% higher diversity than expected to at least ~%50 lower than expected. For D_{XY} this transition goes from 100% higher than expected to ~50% lower than expected. For F_{ST} it goes from ~80% lower than expected to ~30% higher than expected. For Tajima's D it goes from 1 to -1.3. For r^2 it goes from ~30% higher than the neutral value to ~60% the neutral value. For β it goes from ~250% the neutral value to ~50% the neutral value.

A transition between BGS and AOD has been predicted from single population models when looking at the impact of different selective coefficients on diversity and deviation from the neutral site frequency spectrum SFS (Gilbert et al. 2019). As such, if we assume that most deleterious mutations are recessive, which is consistent with mutation accumulation experiments (García-Dorado et al. 1998), then we would expect zones of the genome with deleterious mutations of stronger effect, such as those that occur in conserved region that code for house-keeping genes (She et al. 2009) to be under the effect of BGS, while zones of the genome where mutations have less deleterious effects to be under the effects of AOD. Our results indicate that this transition does not qualitatively depend on the migration rate, but it is simply influenced by the strength of selection. However, this conclusion might depend on demographic and selective factors that were not investigated here. For instance, it would be required to test if this transition is independent of the effects of: (1) demographic history, in particular, different effective sizes in ancestral and descending populations and well as the time of split; (2) variation in the selective coefficients, rather than assuming a fixed selective coefficient; and (3) variation in the dominance coefficient, rather than assuming that all mutations are fully recessive.

5.5 Distinguishing between BGS and AOD in genomic data from natural populations

We found a correlation between various statistics, depending on whether BGS or AOD occurs. When recombination is low, we found that the genomic signature of BGS is low diversity π and low $D_{xy,}$ a negative Tajima's D, high F_{ST} and low R^2 and low β. The genomic signature of AOD is high diversity and high D_{xy}, a positive Tajima's D, low F_{ST}, high r^2 and high β statistic.

Interestingly, we found that regions with high recombination do not show this pattern, exhibiting values close to the neutral expectation. Hence, what we mean by "low" or "high" is in relation to what is observed in regions of the genome with higher recombination rates, as those reflect the neutral expectation due to demographic history. Thus, our results indicate that it might be possible to develop simple statistical tests comparing regions of the genome with different recombination rates, in order to detect BGS and AOD, using regions of high recombination as a benchmark. In order to test these predictions, we would need access to both genomic polymorphism and differentiation (F_{ST}) from pairs of populations that have or are experiencing migration and from which the recombination map in known. If we also have access to data from both populations with different migration rates between them, we can test the prediction that different migration rates lead to similar genomic patterns.

5.6 Can BGS and AOD lead to false positives in genome scans?

The results we found have implications when it comes to the detection of divergent selection and genes involved in local adaptation. These tests rely on genome scans of F_{ST} and diversity and assume that an increased F_{ST} in a region of the genome is a sign of selection in that region while a lowered F_{ST} is a sign of migrations. However, our results show that these two statistics can be affected by linked selection due to deleterious mutations, even with when migration occurs. Because BGS can lead to an increase of F_{ST} with low recombination rates, it can lead to the detection of false positives. Indeed, as reported in several studies and reviewed by (Cruickshank and Hahn 2014), regions with high differentiation (F_{ST}) are usually found in regions of low recombination. This is exactly what is predicted for BGS, and hence our results suggest that BGS can lead to false positives even when there is some migration ($2N_em<10$). However, we note that this effect is reduced (less than a 30% increase in F_{ST}). It would be required to further investigate if regions experiencing the joint effects of BGS and divergent selection lead to higher peaks of differentiation.

The occurrence of AOD, on the other hand, leads to a decrease of F_{ST} in regions of low recombination. This lower F_{ST} in regions of low recombination predicted by the effects of AOD is different from what is expected due to divergent selection, and it is different than the reported in most studies of recently diverged species experiencing gene flow. This suggests that the impact of AOD does not mimic the effects of divergent selection and does not lead to false positives. However, we found that AOD can even affects genomic regions with intermediate recombination, which means it can have a genome-wide effect. This has an indirect implication for genome scans, as usually the statistical methods rely on the global distribution of F_{ST}. Because AOD decreases the mean F_{ST} it can affect demographic parameters used to determine the neutral null model.

One of the possible solutions is to use high recombination zones of the genome as a benchmark due their "neutral like behaviour". This would allow us to calculate the distribution of F_{ST} across the genome due to neutral demographic processes, i.e. to find the "neutral" F_{ST} baseline against which the potential regions of the genome would be compared to.

6 Conclusions

Our results show that the occurrence of purifying linked selection can lead to the occurrence of BGS or AOD if recombination rates are low. These two processes affect various statistics that used to characterize natural population and, as such, can be detected using previously developed tools to study natural populations. The transition between AOD and BGS that occurs due to selection can be seen when measuring various statistics This transition is independent of migration rate. Migration also does not change the genomic patterns we observe, except it removes the increase in F_{ST} seen with BGS if it is very high ($2N_em$=10).

The detection of divergent selection has been based on the genome scans to detect peaks of F_{ST} and valleys of diversity. Our results show that these peaks in F_{ST} may instead be caused by the occurrence of BGS in regions of the genome with low recombination. In order to solve this issue one can used the zones of the genome with high recombination rate as a neutral baseline, against which other regions can be compared to find those with significant outlier peaks of F_{ST}. Another possible solution is to develop methods to detect selection that account and model explicitly the influence that linked selection has on the F_{ST}.

Our results generate several predictions that could be compared against genomic data from natural populations. In the future, in order to test these predictions it would be necessary to obtain genomic data from natural or experimental populations from which there is a known recombination map, in order to test if the patterns we found in our simulations are verified. From a theoretical perspective, it will also be interesting to model and explore other factors such as the interaction between positive selection and purifying selection or the effects of linked selection on sex chromosomes.

7 Bibliography

Agrawal, A. F., and M. C. Whitlock. 2011. Inferences about the distribution of dominance drawn from yeast gene knockout data. Genetics 187:553–566. Genetics Society of America.

Auton, A., A. Fledel-Alon, S. Pfeifer, O. Venn, L. Segurel, T. Street, E. M. Leffler, R. Bowden, I. Aneas, J. Broxholme, P. Humburg, Z. Iqbal, G. Lunter, J. Maller, R. D. Hernandez, C. Melton, A. Venkat, M. A. Nobrega, R. Bontrop, S. Myers, P. Donnelly, M. Przeworski, and G. McVean. 2012. A Fine-Scale Chimpanzee Genetic Map from Population Sequencing. Science (80-.). 336:193–198.

Barton, N. H. 2000. Genetic hitchhiking. Philos. Trans. R. Soc. B Biol. Sci. 355:1553–1562. The Royal Society.

Beaumont, M. A., and R. A. Nichols. 1996. Evaluating loci for use in the genetic analysis of population structure. Proc. R. Soc. B Biol. Sci. 263:1619–1626. The Royal Society.

Bersabé, D., A. Caballero, A. Pérez-Figueroa, and A. García-Dorado. 2016. On the Consequences of Purging and Linkage on Fitness and Genetic Diversity. G3 Genes|Genomes|Genetics 6:171–181.

Bhatia, G., N. Patterson, S. Sankararaman, and A. L. Price. 2013. Estimating and interpreting FST: The impact of rare variants. Genome Res. 23:1514–1521. Cold Spring Harbor Laboratory Press.

Booker, T. R., B. C. Jackson, and P. D. Keightley. 2017. Detecting positive selection in the genome. BMC Biol. 15:98. BMC Biology.

Campos, J. L., and B. Charlesworth. 2019. The Effects on Neutral Variability of Recurrent Selective Sweeps and Background Selection. Genetics 212:287–303.

Campos, J. L., L. Zhao, and B. Charlesworth. 2017. Estimating the parameters of background selection and selective sweeps in *Drosophila* in the presence of gene conversion. Proc. Natl. Acad. Sci. 114:E4762–E4771.

Charlesworth, B., M. Nordborg, and D. Charlesworth. 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genet. Res. 70:155–74.

Chen, J.-M., D. N. Cooper, N. Chuzhanova, C. Férec, and G. P. Patrinos. 2007. Gene conversion: mechanisms, evolution and human disease. Nat. Rev. Genet. 8:762–775. Nature Publishing Group.

Comeron, J. M., R. Ratnappan, and S. Bailin. 2012. The Many Landscapes of Recombination in Drosophila melanogaster. PLoS Genet. 8:e1002905. Public Library of Science.

Cruickshank, T. E., and M. W. Hahn. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol. Ecol. 23:3133–3157.

Cvijović, I., B. H. Good, and M. M. Desai. 2018. The effect of strong purifying selection on genetic diversity. Genetics 209:1235–1278.

Eyre-Walker, A., and P. D. Keightley. 2007. The distribution of fitness effects of new mutations. Nat. Rev. Genet. 8:610–618.

Feder, J. L., S. P. Egan, and P. Nosil. 2012. The genomics of speciation-with-gene-flow. Trends Genet. 28:342–350.

Fry, J. D., and S. V. Nuzhdin. 2003. Dominance of mutations affecting viability in Drosophila melanogaster. Genetics 163:1357–1364.

García-Dorado, A., J. L. Monedero, and C. López-Fanjul. 1998. The mutation rate and the distribution of mutational effects of viability and fitness in Drosophila melanogaster. Genetica 102/103:255–265. Kluwer Academic Publishers.

Gilbert, K. J., F. Pouyet, L. Excoffier, and S. Peischl. 2019. Transition from background selection to associative overdominance promotes diversity in regions of low recombination. 1–22.

Haller, B. C., and P. W. Messer. 2018. SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model. Mol. Biol. Evol. 36:632–637.

Hamilton, M. B. 2009. Population genetics. Wiley-Blackwell.

Hill, W. G., and A. Robertson. 1968. Linkage disequilibrium in finite populations. Theor. Appl. Genet. 38:226–231. Springer-Verlag.

Hoban, S., J. L. Kelley, K. E. Lotterhos, M. F. Antolin, G. Bradburd, D. B. Lowry, M. L. Poss, L. K. Reed, A. Storfer, and M. C. Whitlock. 2016. Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. Am. Nat. 188:379–397. NIH Public Access.

Huang, Y. F., and A. Siepel. 2019. Estimation of allele-specific fitness effects across human protein-coding sequences and implications for disease. Genome Res. 29:1310–1321.

Hudson, R. R., and N. L. Kaplan. 1995. Deleterious background selection with recombination. Genetics 141:1605–1617.

Kirkpatrick, M., and N. Barton. 2006. Chromosome inversions, local adaptation and speciation. Genetics 173:419–434.

Matthey-Doret, R., and M. C. Whitlock. 2019. Background selection and F ST : consequences for detecting local adaptation. Mol. Ecol. mec.15197.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. U. S. A. 70:3321–3.

Nordborg, M. 1997. Structured coalescent processes on different time scales. Genetics 146:1501–1514.

Nordborg, M., B. Charlesworth, and D. Charlesworth. 1996. The effect of recombination on background selection. Genet. Res. 67:159. Cambridge University Press.

Orlando, L., and P. Librado. 2019. Origin and Evolution of Deleterious Mutations in Horses. Genes (Basel). 10:649. Multidisciplinary Digital Publishing Institute.

Paælsson, S. 2004. On the effects of background selection in small populations on comparisons of molecular variation. Hereditas 141:74–80.

Pálsson, S., and P. Pamilo. 1999. The effects of deleterious mutations on linked, neutral variation in small populations. Genetics 153:475–83. Genetics Society of America.

Pamilo, P., and S. Pálsson. 1998. Associative overdominance, heterozygosity and fitness. Heredity (Edinb). 81:381–389.

Phadnis, N., and J. D. Fry. 2005. Widespread correlations between dominance and homozygous effects of mutations: Implications for theories of dominance. Genetics 171:385–392.

Pouyet, F., S. Aeschbacher, A. Thiéry, and L. Excoffier. 2018. Background selection and biased gene conversion affect more than 95% of the human genome and bias demographic inferences. Elife 7:1–21.

R Core Team. 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Schou, M. F., V. Loeschcke, J. Bechsgaard, C. Schlötterer, and T. N. Kristensen. 2017. Unexpected high genetic diversity in small populations suggests maintenance by associative overdominance. Mol. Ecol. 26:6510–6523.

She, X., C. A. Rohl, J. C. Castle, A. V Kulkarni, J. M. Johnson, and R. Chen. 2009. Definition, conservation and epigenetics of housekeeping and tissue-enriched genes. BMC Genomics 10:269. BioMed Central.

Siewert, K. M., and B. F. Voight. 2017. Detecting long-term balancing selection using allele frequency correlation. Mol. Biol. Evol. 34:2996–3005.

Sousa, V., and J. Hey. 2013. Understanding the origin of species with genome-scale data: the role of gene flow. Nat. Rev. Genet. 14:404–414.

Stevison, L. S., K. B. Hoehn, and M. A. F. Noor. 2011. Effects of Inversions on Within- and Between-Species Recombination and Divergence. Genome Biol. Evol. 3:830–841. Narnia.

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

Terhorst, J., J. A. Kamm, and Y. S. Song. 2017. Robust and scalable inference of population history from hundreds of unphased whole genomes. Nat. Genet. 49:303–309. Nature Publishing Group.

Watterson, G. A. 1975. On the Number of Segregating Sites in Genetical Models without Recombination. Theor. Popul. Biol. 276:256–276.

Wilkinson-Herbots, H. M. 2008. The distribution of the coalescence time and the number of pairwise nucleotide differences in the “isolation with migration” model. Theor. Popul. Biol. 73:277–288. Academic Press.

Wolf, J. B. W. W., and H. Ellegren. 2017. Making sense of genomic islands of differentiation in light of speciation. Nature Publishing Group.

Zeng, K., and P. Corcoran. 2015. The effects of background and interference selection on patterns of genetic variation in subdivided populations. Genetics 201:1539–1554.

Zhao, L., and B. Charlesworth. 2016. Resolving the conflict between associative overdominance and background selection.

8 Supplemental Material

The following results are divided in two sections: population with an effective migration rate of 0.05 and population with an effective migration rate of 0.5 or an effective migration rate of 0.5. The y-axis and the x-axis are the same as in the main text. These results show that there is not a qualitative difference between the results of populations experiencing low migration and the results for populations that do not experiencing migration (Figure 4.3 to Figure 4.6).

8.1 Effective migration rate $2N_em$= 0.1

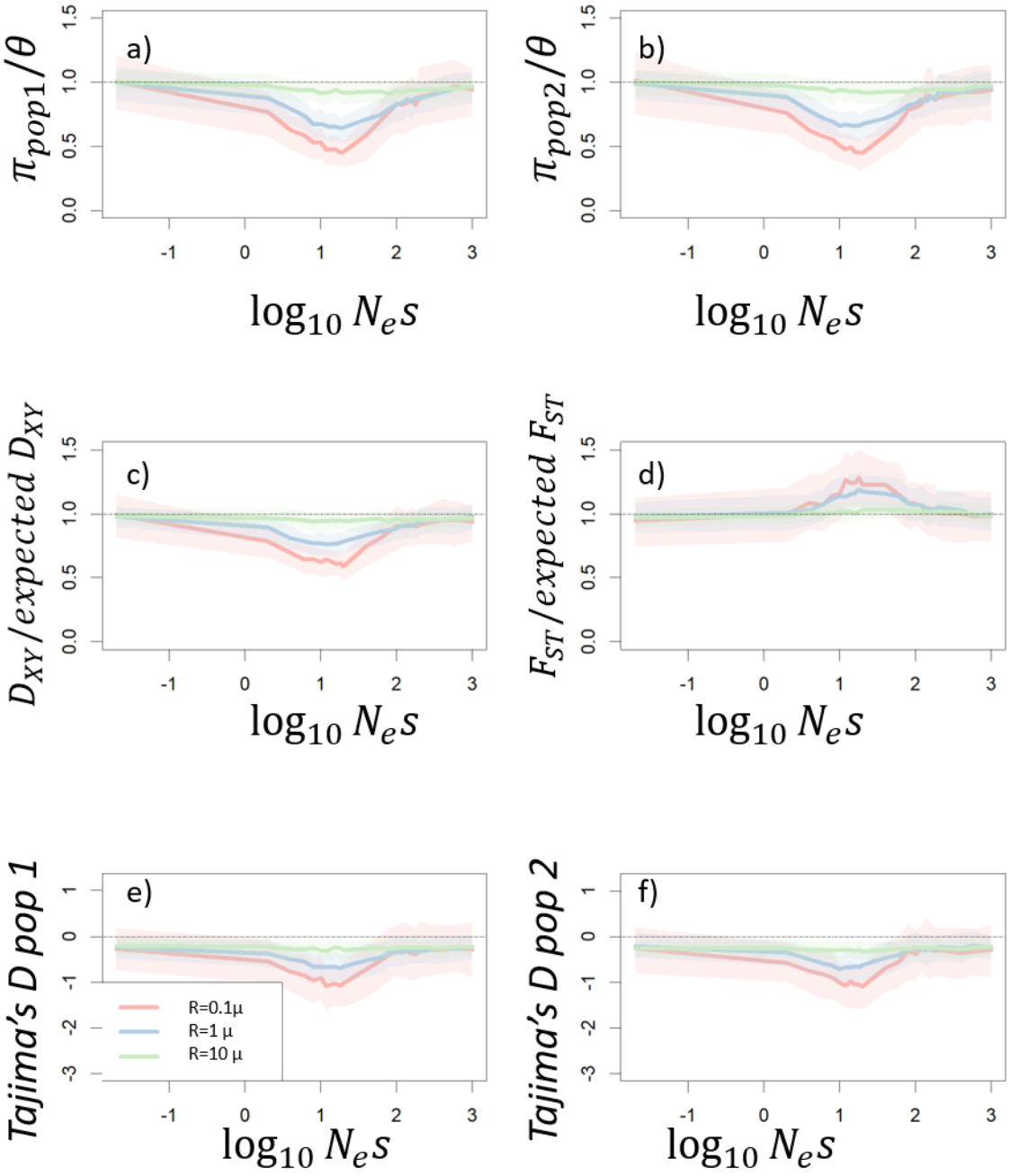

Figure 8.1 **Measures of various statistics for an isolation with migration model with two population and constant migration ($2N_em$=0.1) under the effects of co-dominant deleterious mutations** a) and b) π/(expected π) for population 1 and 2 c) and d) D_{xy} and F_{ST} between both populations d) and e) Tajima's D for population 1 and 2 for neutral mutations. In a)-d) the statistics are shown relative to the neutral "expected" value based on analitycal equatiations, with values equal to 1.0 indicating no deviations from neutrality. Results obtained by performing 200 simulations in SLiM 3.2 for each parameter combination . The solid line represents the mean for the 200 simulations while the shaded area represents the 25% to 75% quantile of the distribution of each statistic across simulations. Different colours correspond to simulations with different recombination rates. Results for neutrality (selection coefficient s=0) are show in log10 scale as -2 as the logarithm of 0 is undefined.

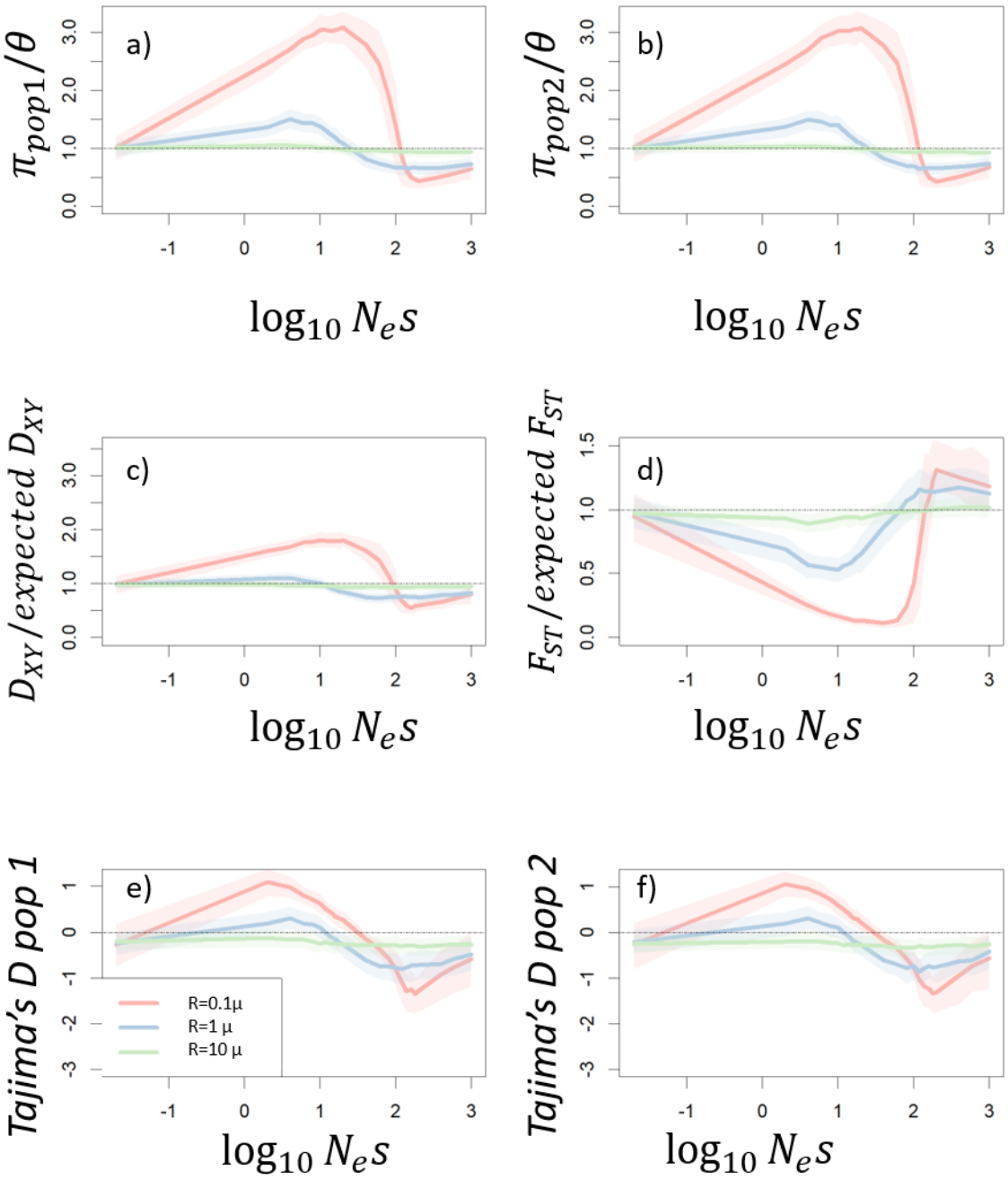

Figure 8.2 **Measures of various statistics for an isolation with migration model with two populations and constant migration ($2N_em=0.1$) under the effects of recessive deleterious mutations.** a) and b) π/(expected π) for population 1 and 2 c) and d) D_{xy} and F_{ST} between both populations, respectively d) and e) Tajima's D for population 1 and 2 for neutral mutations. In a)-d) the statistics are shown relative to the neutral "expected" value based on analitycal equatiations, with values equal to 1.0 indicating no deviations from neutrality. Results obtained by performing 200 simulations in SLiM 3.2 for each parameter combination. The solid lines represents the mean for the 200 simulations while the shaded area represents the 25% to 75% quantile of the distribution of each statistic across simulations. Different colours correspond to simulations with different recombination rates. Results for neutrality (selection coefficient s=0) are show in log10 scale as -2 as the logarithm of 0 is undefined.

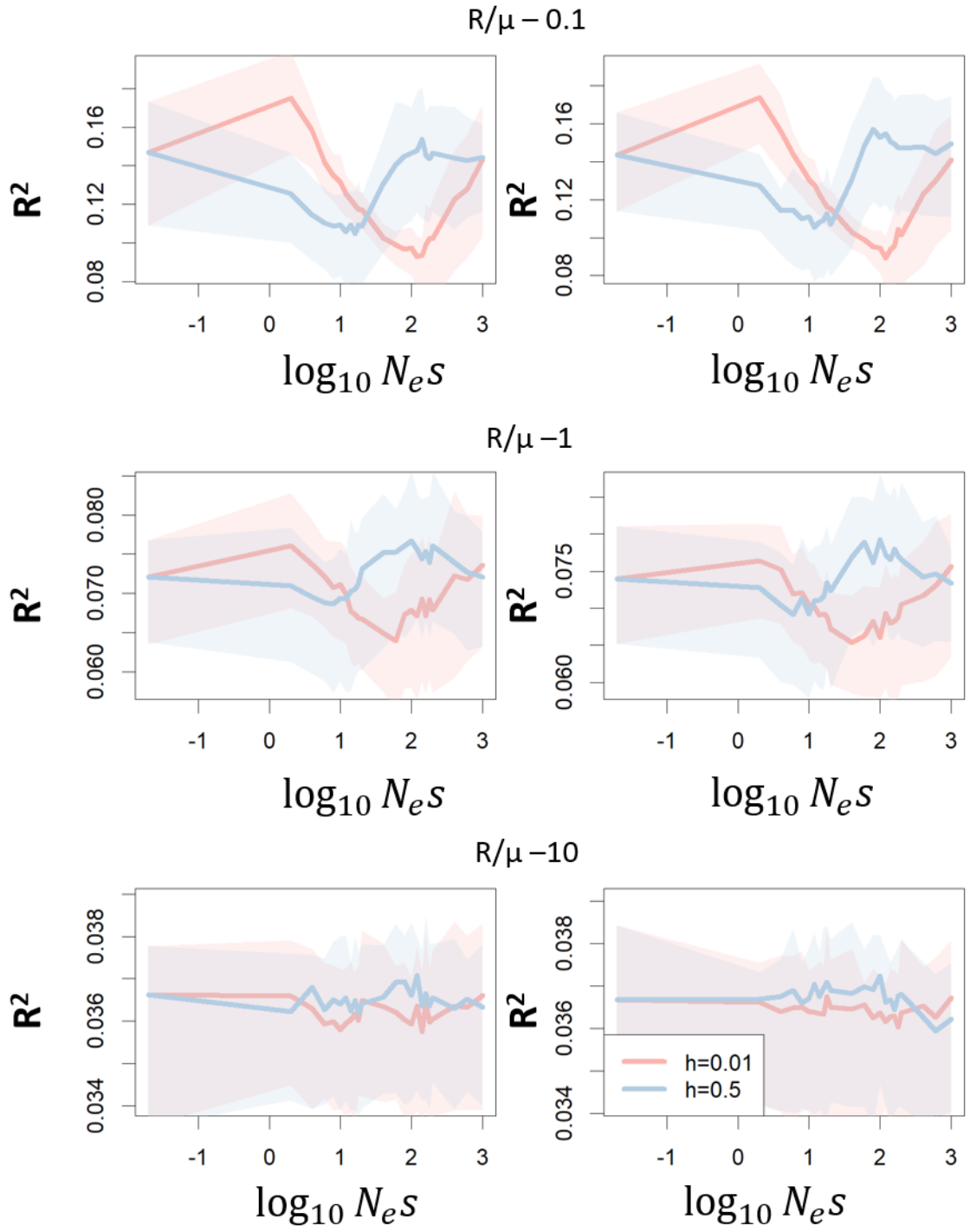

Figure 8.3 ***The effect of linked selection on linkage disequilibrium (r^2) in an isolation with migration model with two constant-size populations of 1,000 individuals and constant migration ($2N_em=0.1$), experiencing removal of co-dominant or recessive deleterious mutations of fixed selective coefficients, with different recombination rates.*** *Solid lines correspond to the mean of 200 simulations performed using SLiM of a 50Kb region with alternating deleterious and neutral sites with a fixed selective coefficient. Note the different scale in the y-axis with each recombination rate. The solid line represents the mean for the 200 simulations while the shaded area represents the 25% to 75% quantile of the distribution of each statistic across simulations. Different colours correspond to simulations with different dominance coefficients h. Results for neutrality (selection coefficient s=0) are show in log10 scale as -2 as the logarithm of 0 is undefined.*

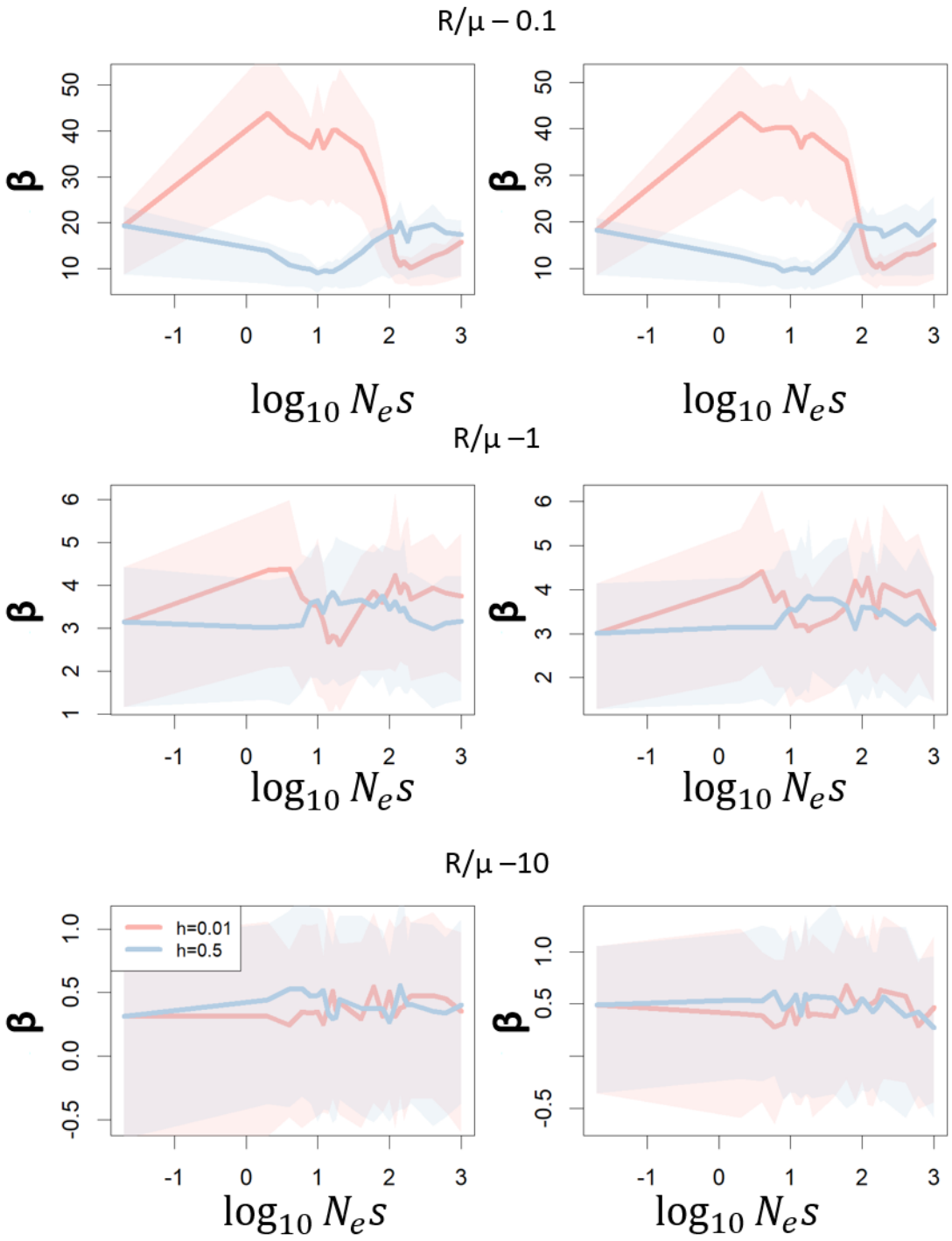

Figure 8.4 ***The effect of linked selection on a statistic to detect balancing selection (β) in an isolation with migration model with two constant-size population of 1,000 individuals and constant migration ($2N_em=1$) experiencing removal of co-dominant or recessive deleterious mutations of fixed selective coefficients, with different recombination rates***. *Solid lines correspond to the mean of 200 simulations performed using SLiM of a 50Kb region with alternating deleterious and neutral sites with a fixed selective coefficient. Note the different scale in the y-axis for each recombination rate. The solid line represents the mean for the 200 simulations while the shaded area represents the 25% to 75% quantile of the distribution of each statistic across simulations. Different colours correspond to simulations with different dominance coefficients h. Results for neutrality (selection coefficient s=0) are show in log10 scale as -2 as the logarithm of 0 is undefined.*

8.2 Effective migration rate $2N_em = 1$

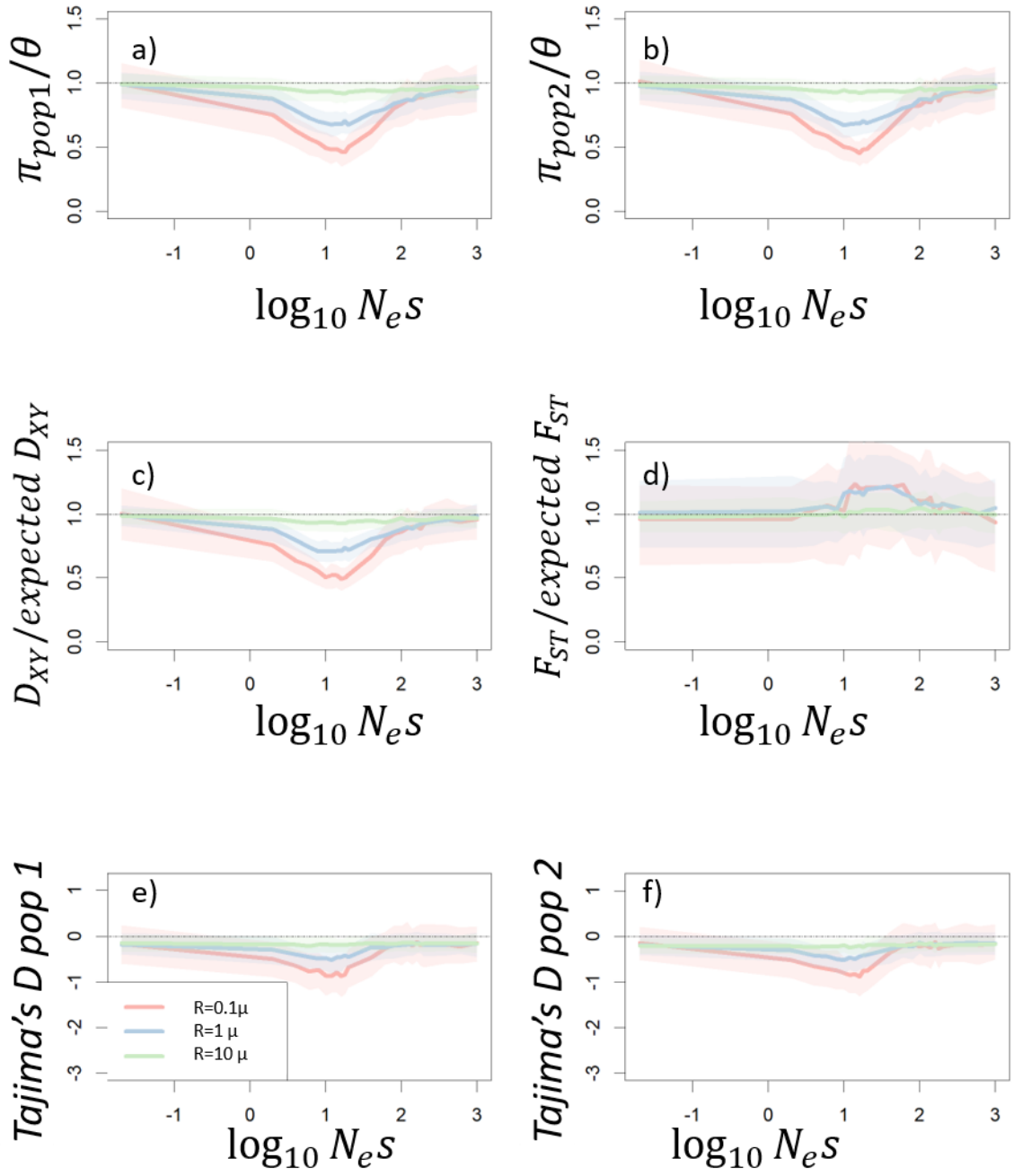

Figure 8.5 **Measures of various statistics for an isolation with migration model with two population and constant migration ($2N_em=1$) under the effects of co-dominant deleterious mutations** a) and b) π/(expected π) for population 1 and 2 c) and d) D_{xy} and F_{ST} between both populations d) and e) Tajima's D for population 1 and 2 for neutral mutations. In a)-d) the statistics are shown relative to the neutral "expected" value based on analitycal equatiations, with values equal to 1.0 indicating no deviations from neutrality. Results obtained by performing 200 simulations in SLiM 3.2 for each parameter combination . The solid line represents the mean for the 200 simulations while the shaded area represents the 25% to 75% quantile of the distribution of each statistic across simulations. Different colours correspond to simulations with different recombination rates. Results for neutrality (selection coefficient s=0) are show in log10 scale as -2 as the logarithm of 0 is undefined.

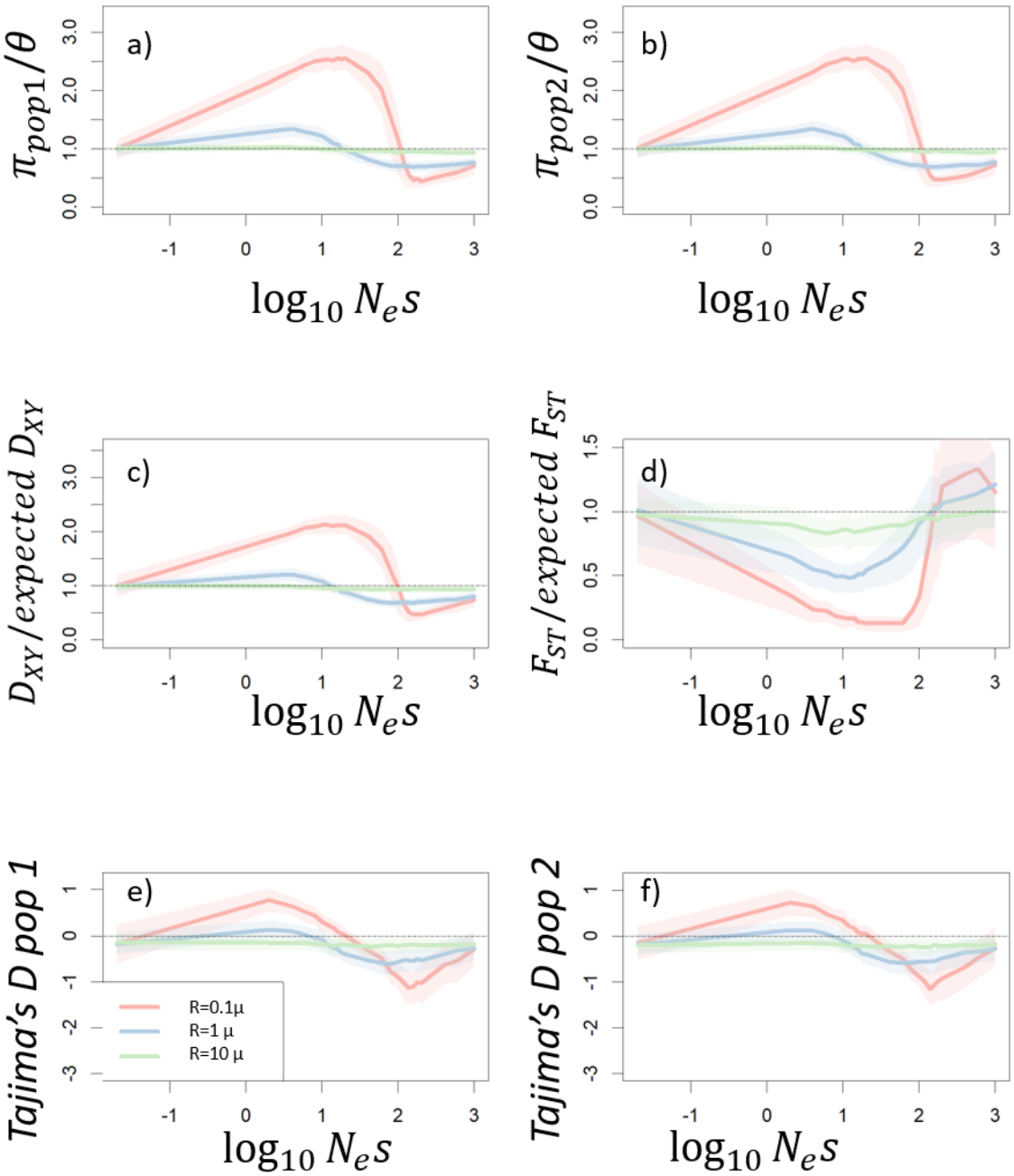

Figure 8.6 ***Measures of various statistics for an isolation with migration model with two populations and constant migration ($2N_em=1$) under the effects of recessive deleterious mutations.*** *a) and b) π/(expected π) for population 1 and 2 c) and d) D_{xy} and F_{ST} between both populations, respectively d) and e) Tajima's D for population 1 and 2 for neutral mutations. In a)-d) the statistics are shown relative to the neutral "expected" value based on analitycal equatiations, with values equal to 1.0 indicating no deviations from neutrality. Results obtained by performing 200 simulations in SLiM 3.2 for each parameter combination. The solid lines represents the mean for the 200 simulations while the shaded area represents the 25% to 75% quantile of the distribution of each statistic across simulations. Different colours correspond to simulations with different recombination rates. Results for neutrality (selection coefficient s=0) are show in log10 scale as -2 as the logarithm of 0 is undefined.*

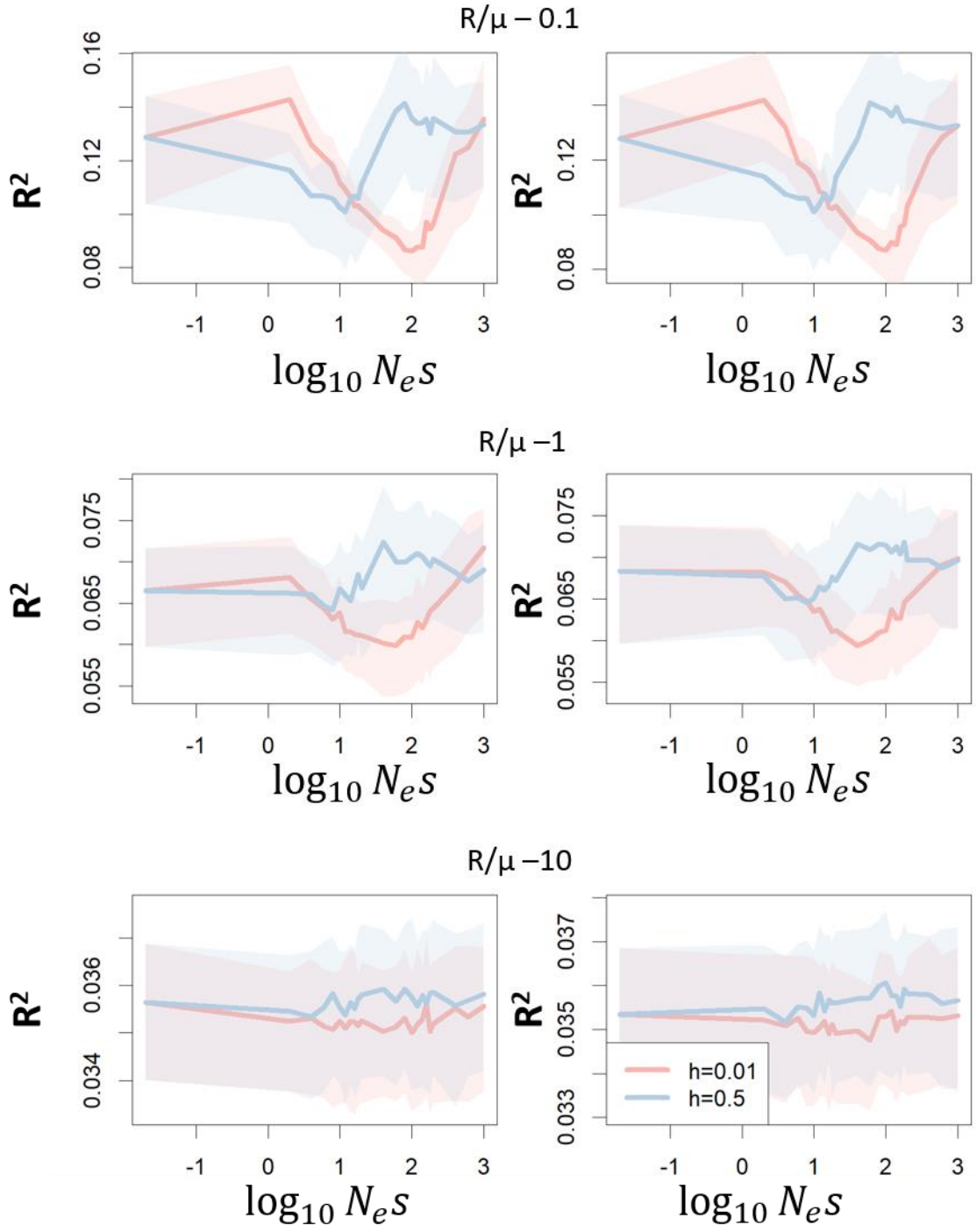

Figure 8.7 ***The effect of linked selection on linkage disequilibrium (r^2) in an isolation with migration model with two constant-size populations of 1,000 individuals and constant migration ($2N_em=1$), experiencing removal of co-dominant or recessive deleterious mutations of fixed selective coefficients, with different recombination rates.*** *Solid lines correspond to the mean of 200 simulations performed using SLiM of a 50Kb region with alternating deleterious and neutral sites with a fixed selective coefficient. Note the different scale in the y-axis with each recombination rate. The solid line represents the mean for the 200 simulations while the shaded area represents the 25% to 75% quantile of the distribution of each statistic across simulations. Different colours correspond to simulations with different dominance coefficients h. Results for neutrality (selection coefficient s=0) are show in log10 scale as -2 as the logarithm of 0 is undefined.*

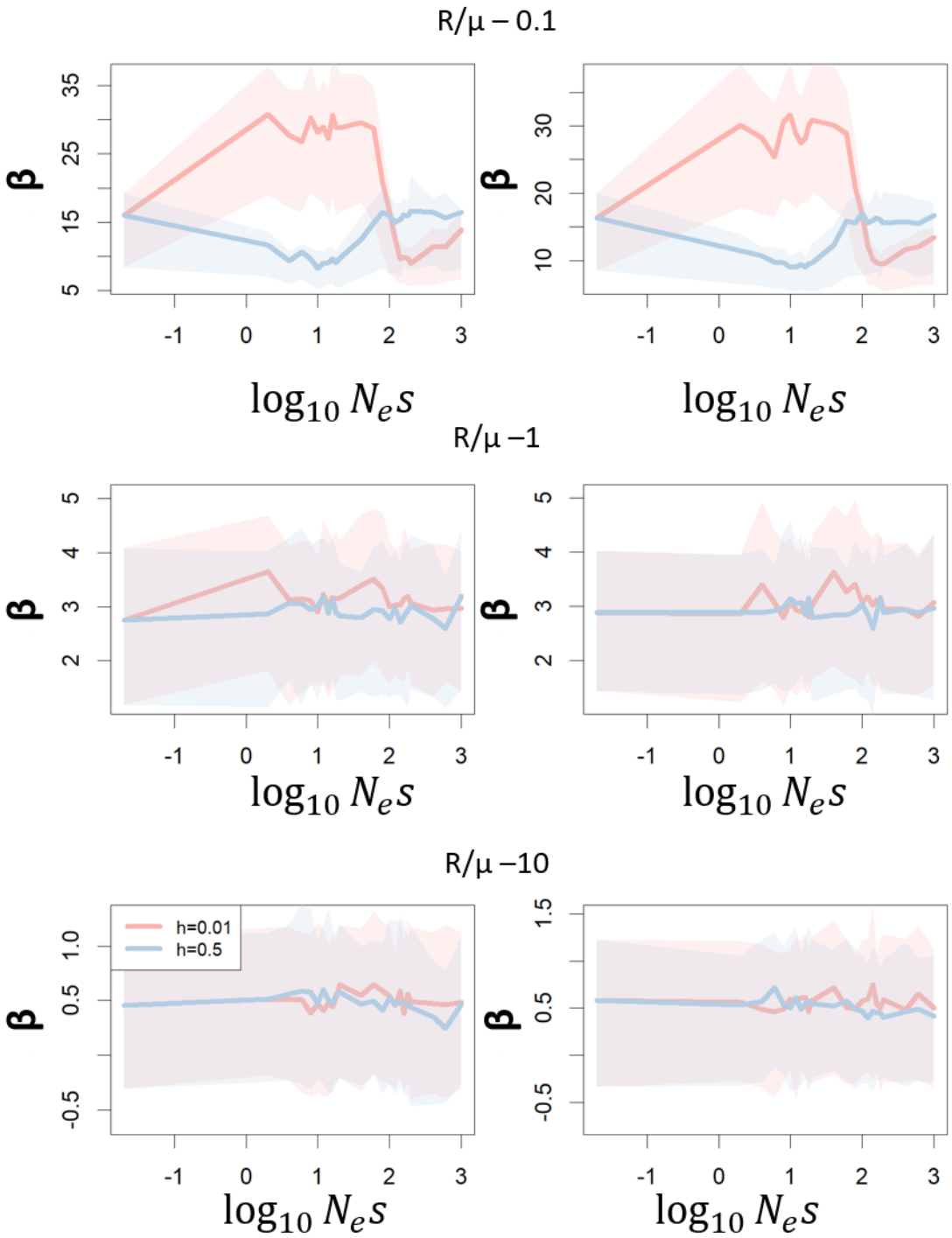

Figure 8.8 ***The effect of linked selection on a statistic to detect balancing selection (β) in an isolation with migration model with two constant-size population of 1,000 individuals and constant migration ($2N_em=1$) experiencing removal of co-dominant or recessive deleterious mutations of fixed selective coefficients, with different recombination rates****. Solid lines correspond to the mean of 200 simulations performed using SLiM of a 50Kb region with alternating deleterious and neutral sites with a fixed selective coefficient. Note the different scale in the y-axis for each recombination rate. The solid line represents the mean for the 200 simulations while the shaded area represents the 25% to 75% quantile of the distribution of each statistic across simulations. Different colours correspond to simulations with different dominance coefficients h. Results for neutrality (selection coefficient s=0) are show in log10 scale as -2 as the logarithm of 0 is undefined.*