

# The Effect of Sexual Selection on Cline Patterns in Biological Traits

Case study of *Littorina saxatilis*

Master's thesis in Complex Adaptive Systems

Niclas Ståhl



MASTER'S THESIS 2016

# The Effect of Sexual Selection on Cline Patterns in Biological Traits

Case study of *Littorina saxatilis*

Niclas Ståhl



**CHALMERS**  
UNIVERSITY OF TECHNOLOGY

Department of Physics  
CHALMERS UNIVERSITY OF TECHNOLOGY  
Gothenburg, Sweden 2016

The Effect of Sexual Selection on Cline Patterns  
Case study of *Littorina saxatilis*  
Niclas Ståhl

© Niclas Ståhl, 2016.

Supervisor: Marina Rafajlovic, Department of Physics  
Examiner: Bernhard Mehlig, Department of Physics

Master's Thesis 2016  
Department of Physics  
Chalmers University of Technology  
SE-412 96 Gothenburg  
Telephone +46 31 772 1000

Cover: Average allele-effect size as a function of deme position. This result is obtained 60000 generations after the start of the simulation of a model where male fecundity depends on size. This model is described in Section 2.4. The red line is the trace of midpoints through time and the midpoint at the current generation is represented with a red star. The expected steady state in the case of no migration and no increased fecundity in males is shown as a blue dashed line. The black dotted line shows the centre of the population where the environmental shift occurs. Parameters:  $\sigma = 8$ ,  $m = 2$  and  $d = 0.1$ ,  $\mu = 0.0002$ ,  $\mu_{sd} = 0.05$ . There are 80 demes in the simulation each with 100 individuals. For more information see the caption of Figure 3.7.

Typeset in L<sup>A</sup>T<sub>E</sub>X  
Gothenburg, Sweden 2016

## Abstract

Natural selection makes it possible for biological populations to adapt to their local environmental conditions. A species living in habitats with contrasting conditions may evolve differences in biological traits in the different habitats. If gene flow exists between two differently adapted populations a biological trait that is under selection may exhibit a gradual change in space. This gradual change in a biological trait over space is commonly referred to as *cline in the trait*. Examples include clines in size or thickness of shell of individuals of the marine snail *Littorina saxatilis*. Empirical studies of clines in different traits in *L. saxatilis* have found that the midpoint of the cline in size is significantly shifted from the geographic position of the environmental change, whereas no significant shift is found in other clines. This is puzzling since typically theoretical studies find that on average the midpoint of a cline coincides with the geographic position of environmental change. The shift occurs in the same direction in three independent populations. Therefore it is unlikely that it is detected by chance. The aim of this Master thesis is to investigate whether such a shift could result from sexual selection. To this end, two models with different types of sexual selection are presented. One model is based on assortative mating, and the other on that the reproductive success of males depends on size. The use of these models is motivated by data from recent laboratory experiments performed within the Linnaeus Centre for Marine Evolutionary Biology, University of Gothenburg (data obtained and analysed with permission from Kerstin Johannesson). In this thesis it is found that assortative mating does not produce a shift of the midpoint of a cline. By contrast, it is shown in this thesis that the model in which the reproductive success of males increases with increasing size gives rise to a shift of the midpoint of the resulting size cline. The shift appears during a transient state which lasts for up to  $10^5$  generations under the parameters tested here. The shift, however, disappears when the system reaches a steady state. Considering that *Littorina saxatilis* colonised its current habitat after the last glacial period, which ended about 10000 years ago, *L. saxatilis* is likely to be in the transient state. If so, a shift caused by sexual selection is expected to be observed today. It remains to be understood whether the midpoint of the clines in other traits is affected by a physical linkage of the genes underlying them to the genes underlying the trait that is under sexual selection.

Keywords: hybrid zone, assortative mating, size dependent fecundity, natural selection, dispersal, genetic drift

## Acknowledgements

I would like to express my great gratitude to my supervisor Marina Rafajlović for needed guidance and helpful discussions. Kerstin Johannesson and Roger K. Butlin have explained all the biological concepts of the fascinating marine snail *Littorina saxatilis*, and I am very grateful for their guidance and optimism. They do also have my gratitude for proposing the main question of this thesis and suggesting possible hypotheses and approaches. Finally I would like to thank my family for their support and help throughout this Master thesis.

Niclas Ståhl, Gothenburg, 2016







# Contents

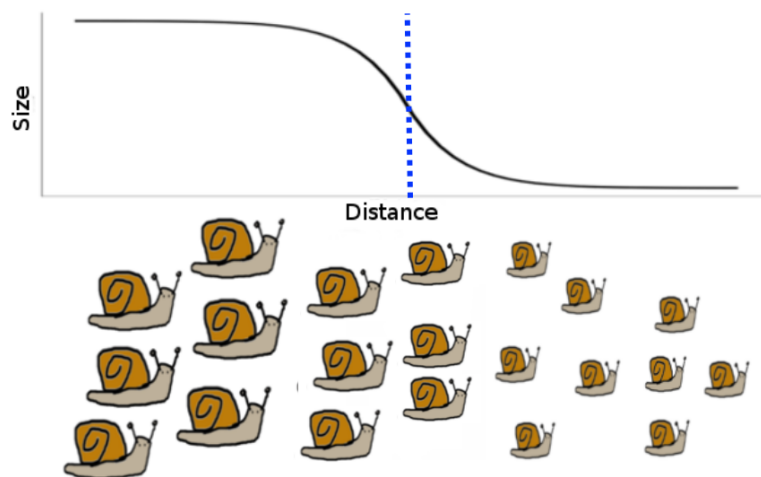
<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Model and Methods</b>	<b>7</b>
2.1	Measuring the slope and midpoint of a cline . . . . .	7
2.2	Assumptions in the models . . . . .	8
2.3	A model with assortative mating . . . . .	9
2.4	A model in which male fecundity depends on size . . . . .	10
<b>3</b>	<b>Results</b>	<b>13</b>
3.1	Comparison with existing theory . . . . .	13
3.2	The effect of assortative mating . . . . .	18
3.3	The effect of size dependent male fecundity . . . . .	21
<b>4</b>	<b>Discussion and Conclusions</b>	<b>27</b>
	<b>Bibliography</b>	<b>31</b>
<b>A</b>	<b>Mathematical derivations</b>	<b>I</b>
A.1	Slope of a cline . . . . .	I
A.2	Analyse of the polymorphic steady state . . . . .	V
A.2.1	Conditions for the existence of a polymorphic steady state . . . . .	V
A.2.2	Stability analysis of the steady states . . . . .	VIII
<b>B</b>	<b>Statistical analysis</b>	<b>XI</b>
B.1	Statistical analysis of data from laboratory trials . . . . .	XI
B.2	Estimating parameters of the model . . . . .	XII



# 1

## Introduction

All living individuals have genes. Genes encode for the properties and behaviours of an individual. The total set of genes in an individual is called the genotype. The genotype in combination with environmental factors determines the set of traits, referred to as the phenotype of the individual. A cline in a biological trait is a gradual spatial change of the average trait value in a population. An illustration of the cline in size of the marine snail *Littorina saxatilis* is shown in Figure 1.1.



**Figure 1.1:** An illustration of the clinal pattern in the size trait in *L. saxatilis*. On the left side of the blue dashed line the population is exposed to crab predation and natural selection favours large individuals. On the right side of the blue dashed line the population is exposed to waves and small individuals are favoured by natural selection. Due to migration there is a region in the middle where large and small individuals meet and mate. This region is called a hybrid zone and is inhabited by intermediate sized snails, (so called *hybrids*).

Clines are frequently observed in natural populations, and they typically indicate that natural selection is at work. Two main hypotheses on how clines are established and maintained have been presented earlier (reviewed by Felsenstein, 1976; Fife, 1979). The first theory suggests that a cline is established and maintained by natural selection towards different optimal phenotypes in different parts of the habitat. The second theory suggests that a cline emerges due to a secondary contact between two species that can interbreed. This type of clines are maintained by selection against hybrids. The reason why selection acts against hybrids is that the

two interbreeding species have been separated from each other for a long time and accumulated different mutations. Some mutations that are accumulated in one of the species are likely to be incompatible with some mutations in the other species. When the individuals from the different species breed, incompatible mutations are combined and the fitness of the offspring is, therefore, low (Johnson, 2008).

Both these hypotheses stem from mathematical theories that build on the work of Fisher (1937). The first theory originates from the works of Haldane (1948) and Slatkin (1973) while the second theory is based on the work by Bazykin (1969). Fisher (1937) proposes a diffusion model of selection and migration in a population, to investigate how an advantageous allele spreads in the population. The solutions to the diffusion model have the form of travelling waves. To explain the behaviour of clines, Haldane (1948) extends the diffusion model used by Fisher (1937) and adds spatial dependent selection, where the optimal trait value abruptly changes in the centre of the habitat. With this model Haldane (1948) study how different selective pressures on the two different sides of the centre affect the shape and slope of the cline that is formed under these conditions. Furthermore Slatkin (1973) continues on the work of Haldane (1948) and generalised the diffusion model. In the model that Slatkin (1973) presents, individuals are diploid and have one locus. The locus can carry one of the two alleles  $A$  or  $a$ . With this model Slatkin (1973) shows that there is a solution to the diffusion equation if the selection pressure is the same on both sides of the centre of the habitat. This model is described in further detail in appendix A.1. Slatkin (1973) finds that the steady state frequency of allele  $A$  at position  $x$  is given by

$$p(x) = \begin{cases} \frac{1}{2} \left( -1 + 3 \tanh^2 \left( \sqrt{\frac{s}{2l}} x + \operatorname{arctanh} \left( \sqrt{2/3} \right) \right) \right), & \text{when } x > 0, \\ \frac{1}{2} \left( 3 - 3 \tanh^2 \left( \sqrt{-\frac{s}{2l}} x + \operatorname{arctanh} \left( \sqrt{2/3} \right) \right) \right), & \text{when } x < 0. \end{cases} \quad (1.1)$$

Here  $l$  is the average squared dispersal distance and  $s$  is the selection parameter that controls the strength of the natural selection.

In contrast to the work of Haldane (1948) and Slatkin (1973), Bazykin (1969) has studied a model of a cline with low hybrid fitness. Bazykin (1969) starts from the same diffusion equation as Fisher (1937) but instead of assuming that one allele is advantageous, (as Fisher (1937) does), Bazykin (1969) assume two alleles, denoted  $A$  and  $a$ , that are incompatible with each other so that heterozygotes are less fit than homozygotes. Bazykin (1969) found that this model has five steady states, three of which are stable. The two unstable steady states have the form of the tangens hyperbolicus function. Denoting the steady state frequency of allele  $A$  at position  $x$  by  $q(x)$ , the five steady states and their stability are as follows Bazykin (1969)

$$q(x) = \begin{cases} 1, & \text{stable state,} \\ \frac{1}{2}, & \text{stable state,} \\ 0, & \text{stable state,} \\ \frac{1}{2} \left( 1 + \tanh \left( \frac{\sqrt{s}(x+\xi)}{\sqrt{l}} \right) \right), & \text{unstable state,} \\ \frac{1}{2} \left( 1 - \tanh \left( \frac{\sqrt{s}(x+\xi)}{\sqrt{l}} \right) \right), & \text{unstable state.} \end{cases} \quad (1.2)$$

Here  $l$  is the average squared dispersal distance,  $s$  is the selection parameter that controls the strength of the natural selection, respectively as explained above and  $\xi$  is an arbitrary constant that determines the midpoint of the cline.

While both hypothesis for the formation and maintenance of clines (described above) are relevant for biological populations, this thesis focuses on clines that form and are maintained by selection towards different optimal phenotypes, because this is a likely scenario in *L. saxatilis* in Sweden. However it is important to be aware of the second hypothesis and the results of Bazykin (1969), since most clines in nature are believed to have formed due to secondary contact and maintained by low hybrid fitness (Barton and Hewitt, 1989). Furthermore, the tangens hyperbolicus function is also often used in empirical studies, (due to its simplicity,) when parameters of clines are estimated (Fitzpatrick, 2013). Empirical studies often consider the slope at the midpoint of the cline and not the shape of the cline (Barton and Hewitt, 1989). At the midpoint where  $x \ll 1$  the slope of the two solutions presented by Slatkin (1973) and Bazykin (1969) is proportional to each other. Finally, note that the solution derived by Slatkin (1973) (see Equation (1.1)) and the solution derived by Bazykin (1969) (see Equation (1.2)) are qualitatively similar (Slatkin, 1973; May et al., 1975). Therefore, clines emerging under the two scenarios are very difficult to distinguish using empirical data.

As mentioned above, an example species where cline patterns are observed is the marine snail *Littorina saxatilis*. In this species cline patterns are formed between two different morphs which lives in two different types of habitats. One of the morphs lives in sheltered bays with high crab predation while the other morph lives on wave exposed rocks (Janson, 1983). By removing some individuals from their native habitat and placing them in the habitat of the other morph, Janson (1983) shows that native individuals have a larger chance of survival than individuals from the opposite habitat. Thus it is likely that selection towards different optimal phenotypes is at work in this populations. Indeed, for protection against crab attacks it is beneficial to be large, have a thick shell and a small aperture (Janson, 1983). Thus in crab-exposed habitats natural selection favours snails with these traits. The opposite is true for the other morph that lives on wave exposed rocks where crab predation is very rare. Here it is beneficial to be small, have a thin shell and a large aperture, since these traits minimise the risk of getting swept out to sea and drown (Janson, 1983). In the junction of these two habitats the two morphs meet and interbreed and hybrids with intermediate traits form. In these hybrid zones,

no significant differences in the likelihood of survival between hybrids and the two different morphs has been observed in *L. saxatilis* (Janson, 1983).

Empirical studies of clines in, for example size, shell thickness and size of aperture have been performed within the Linnaeus Centre for Marine Evolutionary Biology. In these studies it is found that the midpoint of the cline in size does not coincide with the midpoint of the clines in the two other traits (Johannesson and Butlin, personal communication). In addition the the latter clines coincides with the centre of the environmental change between the two habitats. The cline in size is found to be shifted in three populations at different locations on the Swedish west coast. All three shifts found are of a similar magnitude and to the same direction. The populations sampled are far from each other and gene flow between them is very rare (Johannesson and Warmoes, 1990). They are, therefore, essentially independent of each other. This suggests that it is implausible that these shifts are observed by chance. The occurrence of a shift of the cline in size is rather puzzling and intriguing, but there are particular conditions under which a shift occurs. These conditions are briefly explained next.

Firstly, Haldane (1948) propose a model where the optimal phenotype abruptly changes in the centre of a habitat. With this model Haldane (1948) show that if the selective pressure is not the same on the two sides of the centre, the cline is shifted towards the side with weaker selection. Secondly, in a model presented by May et al. (1975) the migration is much greater in a specific direction. This causes the cline to be shifted in the direction where the migration is larger. This model applies to multiple marine species that have a larvae stage where migration tends to be larger in the same direction as the sea currents (Galindo et al., 2010). However *L. saxatilis* does not have a larvae stage. Thirdly, Nagylaki (1978) presented a model where the mobility of individuals is phenotype dependent. In this model individuals with a high mobility spread at a higher rate than the less mobile phenotype causing the cline to be shifted towards the habitat with the less mobile phenotype. Notably, both the model with a higher migration in a specific direction (May et al., 1975) and the model in which the mobility of individuals depends on their phenotype (Nagylaki, 1978) gives rise to an observable shift, with respect to the geographical midpoint, in all all clines. However, this is not the case in *L. saxatilis* where only the size cline is shifted. Finally, genetic drift is a process that may cause a cline to be shifted. Indeed, Felsenstein (1975) and Slatkin and Maruyama (1975) showed that genetic drift shifts a cline back and forth around the average midpoint and distorts its shape. The average cline is wider than a cline without genetic drift, while every particular cline that is affected by genetic drift is narrowed (Polechová and Barton, 2011). A shift, in the same direction and similar magnitude, of the cline in size is observed in three independent populations (Johannesson and Butlin, personal communication). It is therefore unlikely that the shift solely is a consequence of random genetic drift.

However none of the studies mentioned above analyses how mechanisms beyond natural selection and dispersal influence cline patterns in natural populations. This is important to understand since sexual selection are at work in biological populations. This is explained below for the particular example species that this thesis focuses on, namely for *L. saxatilis* (Johannesson et al., 1995; Erlandsson and Rolán-Alvarez,

1998; Hintz Saltin et al., in press; Johannesson and Butlin, personal communication).

Sexual selection is a type of selection where the reproductive or mating success of an individual depends on its traits. Sexual selection can often be detected as a deviation from random mating. Many studies have observed a deviation from random mating in *L. saxatilis* (Johannesson et al., 1995; Erlandsson and Rolán-Alvarez, 1998). In these studies it is suggested that sexual selection is related to size. Indeed, the likelihood that a male mates with a given female is larger if the female is slightly larger than the male (Johannesson and Butlin, personal communication). In this thesis this type of sexual preference is referred to as *assortative mating*. To simplify discussion, in this thesis assortative mating is interpreted as a type of sexual preference (as in Lande (1981) and Kimura et al. (2015), but see Erlandsson and Rolán-Alvarez (1998) for a different interpretation). Importantly, this does not affect the results obtained under the assortative-mating model (see Chapters 2, 3). Note that analysing a model for assortative mating Sadedin et al. (2009) showed that it is more likely for a species to evolve into different morphs if assortative mating is at work. Apart for assortative mating, Hintz Saltin et al. (in press) recently showed that another type of sexual selection may be present in *L. saxatilis*, that is that larger males give rise to more progeny than smaller males. Therefore to explain the cline patterns emerging in populations of *L. saxatilis* it is necessary to understand the effects of of sexual selection. The aim of this Master thesis is thus to examine how the shape of the cline is affected by assortative mating and size dependent fecundity in males. This is addressed by creating two separate models, the first with assortative mating (see Section 2.3) and the second with size dependent fecundity of males (see Section 2.4). These models is then studied with simulations and the result from these simulations are shown in Chapter 3. These results are then interpreted and discussed in Chapter 4.





# 2

## Model and Methods

In this chapter, the models and methods used in this thesis are presented. In Section 2.1 it is explained how the the midpoint and the slope at the midpoint of simulated clines are measured. Then the two models that are used in this thesis, and differ by the type of sexual selection considers, are presented. Before these models are presented in detail the steps that they both have in common are described, Section 2.2. The two types of sexual selection modelled have been shown to be present in *L. saxatilis*. One model assumes assortative mating, Section 2.3, and the other assumes size dependent fecundity in males, Section 2.4.

### 2.1 Measuring the slope and midpoint of a cline

One property of clines that is often examined in empirical studies is the slope at the midpoint of the cline (Barton and Hewitt, 1985). Due to stochastic effects and because the habitat is divided into discrete patches, I have decided to adopt the following criteria for the midpoint of a cline.

- The change of average phenotype between two demes is assumed to be linear.
- The midpoint of the cline is approximated as the position in the habitat where the average phenotype is exactly between the maximal and the minimal value of the local average phenotypes observed along the habitat.
- If multiple points fulfil this condition, the point at the smallest distance to the environmental change is assumed to be the midpoint.

Slatkin (1973) and Slatkin and Maruyama (1975) derived an expression for the slope  $\left. \frac{dp}{dx} \right|_{x \rightarrow 0}$  at the midpoint  $x \rightarrow 0$  of a cline in a continuously distributed population in the limits of weak selection ( $s \ll 1$ ) and weak dispersal ( $l \ll 1$ ), (see Appendix A.1 for a full derivation):

$$\left. \frac{dp}{dx} \right|_{x \rightarrow 0} = \sqrt{\frac{s}{3l}}. \quad (2.1)$$

Here  $p(x)$  is the solution to the diffusion model presented by Slatkin (1973) (the expression in Equation (2.1) is the derivative of Equation (1.1) that is presented in

the introduction),  $l$  is the average squared dispersal distance and  $s$  is the selection parameter. Contrary to the population in the derivation of the slope by Slatkin (1973) that is continuously distributed the population in the simulations performed within this thesis is divided into a finite number of discrete patches with unit distance between them. In a discrete population, which is divided into demes, there is always a finite distance between a given pair of demes. To be able to go to the limit of  $x \rightarrow 0$  in the discrete population the distance between two neighbouring demes must be negligible in comparison with all other length scales of the system. One other length scales in the system is the width of the cline, which is proportional to the inverse slope. The width of the cline is thus proportional to  $\sqrt{\frac{l}{s}}$ . So for the theory to be valid in a population that is discretely distributed the condition  $\frac{l}{s} \gg 1$  and thus selection must be smaller than dispersion ( $s \ll l$ ), (but not too small).

## 2.2 Assumptions in the models

The two models used in this thesis are based on the following assumptions.

- A population is assumed to inhabit  $M$  patches (also referred to as demes). These demes are ordered in a one dimensional array. For simplicity it is assumed that  $M$  is an even number.
- The number of individuals, denoted  $N$ , is assumed to be equal in all demes at the start of each generation.
- Individuals are diploid and assumed to have one locus under selection. The phenotype that is under selection of an individual is solely determined by the genotype at this locus. The phenotype of the individual  $i$  is denoted as  $z_i$ .
- Each allele is assigned an allele-effect size by which it contributes additively to the phenotype.
- The environmental conditions vary in the habitat. Between deme  $\frac{M}{2}$  and deme  $\frac{M}{2} + 1$  there is an abrupt change of the environmental conditions. The optimal phenotype on the left side of this change differs from the optimal phenotype on the right side.
- The optimal phenotype of deme  $k$  ( $k = 1 \dots M$ ) is denoted by  $\theta_k$ . In all simulations  $\theta_k$  is set to

$$\theta_k = \begin{cases} 2, & \text{when } k < \frac{M+1}{2}, \\ -2, & \text{when } k > \frac{M+1}{2}. \end{cases} \quad (2.2)$$

- Generations are assumed to be discrete and non-overlapping.
- In each generation, it is assumed that the following process occurs:

1. birth and mutations,

2. maturation and migration,
  3. selection,
  4. sexual selection, mating and the production of offspring,
  5. death and beginning of the next generation.
- All individuals are assigned a sex at birth. The sex is assumed to affect only the sexual selection but not affect migration or natural selection.
  - Mutations occur after birth at a rate of  $\mu$  per allele, individual and generation. The effect size of a mutation is assumed to be Gaussian distributed with mean 0 and a standard deviation  $\mu_{sd}$ .
  - Locally within each deme the number of surviving offspring to maternity is assumed to be  $N$ . The chance to survive to maternity is assumed to be independent of phenotype and location of an individual.
  - Individuals may migrate from the habitat they are born in. Two different migration patterns are used throughout this thesis. Firstly, individuals are allowed to migrate only to nearest neighbouring demes. The probability to migrate to the left neighbour is equal to the probability to migrate to the right neighbour and the total probability to migrate is denoted by  $m$ . Secondly, the migration distance is assumed to be Gaussian distributed with mean 0 and a variance  $l$ . Individuals that migrate over the boundaries, (to the left of deme 1 and to the right of deme  $M$ ), are pushed back to their native deme. This is to keep the population size of all demes of roughly equal size.
  - Natural selection is assumed to occur after migration. The natural selection considered is viability selection, where the probability for an individual to survive to mate is proportional to its fitness. The fitness of individual  $i$ , in deme  $k$  is denoted by  $\omega_{i,k}$  and is assumed to be determined by the phenotype and the location of individual  $i$

$$\omega_{i,k} = e^{-\frac{(z_i - \theta_k)^2}{2\sigma^2}}, \quad (2.3)$$

where  $\sigma$  is the selection parameter that governs the strength of selection. Selection is stronger the smaller the value of  $\sigma$  gets. This is the same fitness function as used by Sadedin et al. (2009).

- Sexual selection determines the reproductive success of individuals. This step differs between the two models presented in this thesis. In all models it is assumed that individuals mate an infinite number of times. The two types of sexual selection are described in Section 2.3 and Section 2.4.

## 2.3 A model with assortative mating

In this section it is described how assortative mating (one type of sexual selection) is modelled. The model for assortative mating that is described in detail below, is

similar to the model analysed by Sadedin et al. (2009). Both stochastic simulations with a population size of  $N = 100$  and deterministic simulations with the population size  $N \rightarrow \infty$  are conducted with this model. As mentioned in the introduction of the thesis this model is tailored to mimic empirical observations (Johannesson and Butlin, personal communication).

In the model analysed here, the probability that two individuals of given sizes mate is referred to as the attractiveness between the two individuals. A stronger attractiveness between two individuals means that the probability for them to mate is higher. In this thesis, it is assumed that the attractiveness  $\psi_{i,j,k}$  between a male with size  $z_i^{(male)}$  and a female with size  $z_j^{(female)}$  in deme  $k$  is described by

$$\psi_{i,j,k} = e^{-\frac{(z_{i,k}^{(male)} - z_{j,k}^{(female)} + c)^2}{2\sigma_s^2}}, \quad (2.4)$$

where  $c$  is the optimal difference in size between a male and female, and  $\sigma_s$  is the sexual selection parameter that sets the strength of the mating preference. If  $c$  is positive, males prefer females that are larger than themselves, and vice versa for negative values of  $c$ . For  $c = 0$  both males and females prefer to mate with individuals of the same size as themselves.

The probability  $prob_A(i, j, k)$  that a given male and female both survive the natural selection in deme  $k$  and mate with each other is obtained by combining Equation (2.3) and Equation (2.4):

$$prob_A(i, j, k) = \frac{\omega_{i,k}^{(male)} \omega_{j,k}^{(female)} \psi_{i,j}}{\sum_{a=1}^{N_k^{(m)}} \sum_{b=1}^{N_k^{(f)}} \omega_{a,k}^{(male)} \omega_{b,k}^{(female)} \psi_{a,b}}, \quad (2.5)$$

where  $N_k^{(m)}$  is the number of males and  $N_k^{(f)}$  is the number of females in deme  $k$ . Equation (2.5) gives the contribution of the given male and female pair to the pool of offspring in the next generation.

In the next section, a model for the other form of sexual selection, that this thesis analysis, is described.

## 2.4 A model in which male fecundity depends on size

In this section a model in which male fecundity depends on size is explained. This model is motivated by recent findings by Hintz Saltin et al. (in press) that shows that larger males produce more offspring than smaller males. The laboratory experiment conducted by Hintz Saltin et al. (in press) is described in Appendix B. Similar but slightly different statistical analyses of the data found by Hintz Saltin et al. (in press) are performed in this thesis (see Appendix B). The statistical

analyses use  $\chi^2$ -tests and Pearson correlation-tests to show that it is likely that male fecundity increases with size. In Appendix B parameter values are estimated from the empirical data using maximal likelihood.

In the model where male fecundity increases with size the fecundity,  $\phi_i$  of male  $i$ , is assumed to depend on the size ( $z_i$ ) of this male:

$$\phi_i = 1 - d \left( \frac{\theta_1 - z_i}{\theta_1 - \theta_M} \right). \quad (2.6)$$

Here  $\theta_1$  is the optimal phenotype in deme 1,  $\theta_M$  is the optimal phenotype in deme  $M$  and  $d$  is the male-size fecundity parameter that governs how much the fecundity is affected by the size of the male. The parameter  $d$  is assumed to be in the range  $-1$  to  $1$  where  $-1$ , and  $1$  are not included. A negative  $d$  value means that fecundity decreases with increasing size and a positive  $d$  value means that the fecundity increases with increasing size. When  $d = 0$  the fecundity is not affected by size. This thesis considers positive values of  $d$  but a similar analysis can be performed on negative values of  $d$ . A estimated value of  $d$  in *L. saxatilis* is estimated in Appendix B. The probability  $prob_F(i, j, k)$  that a given male  $i$  and a female  $j$  in deme  $k$  both survive natural selection and mate with each other is obtained by combining Equation (2.3) and Equation (2.6):

$$prob_F(i, j, k) = \frac{\omega_{j,k}^{(female)}}{\sum_{b=1}^{N_k^{(f)}} \omega_{b,k}^{(female)}} \frac{\omega_{i,k}^{(male)} \phi_i}{\sum_{a=1}^{N_k^{(m)}} \omega_{a,k}^{(male)} \phi_i}. \quad (2.7)$$

Here  $N_k^{(m)}$  is the number of males and  $N_k^{(f)}$  is the number of females in deme  $k$ . This is effectively the contribution to the pool of offspring in the next generation of the given male and female pair.

Both models presented above reduce to the model used by Slatkin (1973) when the following hold true:

- Selection is weak,  $\sigma$  is large.
- No mutation,  $\mu = 0$ .
- There are two different alleles present in the population, allele  $A$  with allele-effect size  $\frac{\theta_1}{2}$  and allele  $a$  with allele-effect size  $\frac{\theta_M}{2}$ .
- No assortative mating,  $\sigma_s \rightarrow \infty$ .
- No increased fecundity of males with increasing size,  $d = 0$ .

The model with assortative mating can also be reduced to the model presented by Sadedin et al. (2009) in the special case where  $c = 0$  and  $\sigma_s^2$  from the model presented here is equal to  $\frac{\sigma_a^2}{(2c-1)^2}$  in the model by Sadedin et al. (2009). All males in the model

by Sadedin et al. (2009) must have the same trait value  $c$  and this trait value must be larger than 0.5 as well.

Note that in the plots shown in the next chapter, deme  $k$   $k = (1, \dots, M)$  is assigned a position in the habitat, such that  $x_k = k - \frac{M+1}{2}$ . Using these notations, the environment change in the habitat occurs at the origin. Finally, in the cases where deterministic simulations are presented and shown the simulations are run until the steady state is reached. The system is considered to be in a steady state if the total change in allele frequency in all demes is less than  $10^{-8}$ .

# 3

## Results

This chapter describes the simulation results obtained for the models described in Chapter 2. This chapter is divided into three different sections. In Section 3.1 the analytical results for the slope of a cline derived by Slatkin (1973) is compared to the slopes obtained using simulations. The effect of genetic drift on shape of a cline is also shown. Section 3.2 shows how assortative mating affects the shape of a cline. Finally Section 3.3 shows the results from the model in which male fecundity is determined by the size of the male.

### 3.1 Comparison with existing theory

This section shows a comparison between analytical results and the simulated results from the model used by Slatkin (1973) (described in Appendix A.1). Both models presented in Chapter 2 can be reduced to this model in certain limits (see previous chapter).

In this model no mutations occur and there are only two alleles present in the population, allele  $A$  and  $a$ . It is furthermore assumed that allele  $A$  is favoured on the left side of the environmental change and allele  $a$  is favoured on the right side. Simulation results of how the frequency of allele  $A$  vary over space are depicted in Figure 3.1. This figure shows the results for six different parameter settings. In panels shown on the left side in Figure 3.1 selection is stronger than the corresponding panel on the right side. In the pair of panels on each row the fraction between the migration rate,  $m$ , and the selection parameter,  $s$  is equal. According to the theory by Slatkin (1973) the slope of the clines should therefore be the same in panels in the same row.

The slope of a cline obtained using simulations agrees well with the corresponding theory derived by Slatkin (1973). However when the three conditions, weak dispersal ( $m \ll 1$ ), weak selection ( $s \ll 1$ ) and  $m$  sufficiently larger than  $s$  are not fulfilled the theory fails (see Figure 3.2). This result agrees with the findings of Mallet et al. (1990). Figure 3.2 shows the slope of many deterministic simulations with a infinitely large populations in comparison with the expected slope given by the theory derived by Slatkin (1973). The theoretical slope is presented in Equation (2.1). In Figure 3.2 six different values of the fraction  $\frac{s}{m}$  are selected. The reason for keeping the fraction  $\frac{s}{m}$  constant is that the theoretical slope is fixed when  $\frac{s}{m}$  is constant (see Equation (2.1)). For all simulations conducted the selection parameter

### 3. Results

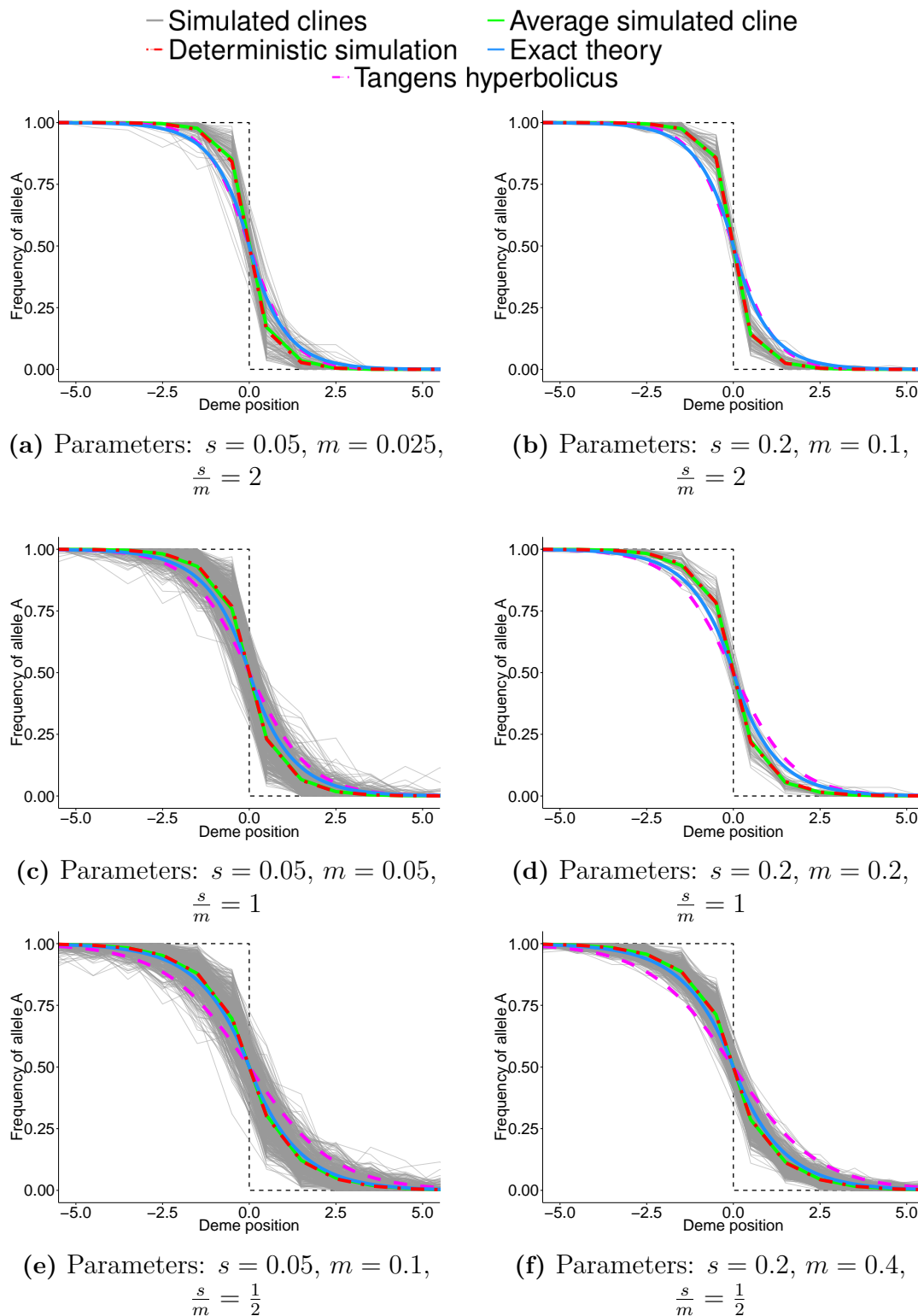
---

$s$  takes values from  $10^{-4}$  to  $10^{-2}$ . For a given value of  $s$  the parameter  $m$  is chosen so that  $\frac{s}{m}$  is constant. To keep simulations biologically relevant, all simulations where the migration rate  $m$  is larger than 0.5 are neglected. This is the reason for why fewer simulations are conducted for smaller values of the fraction  $\frac{s}{m}$  in Figure 3.2.

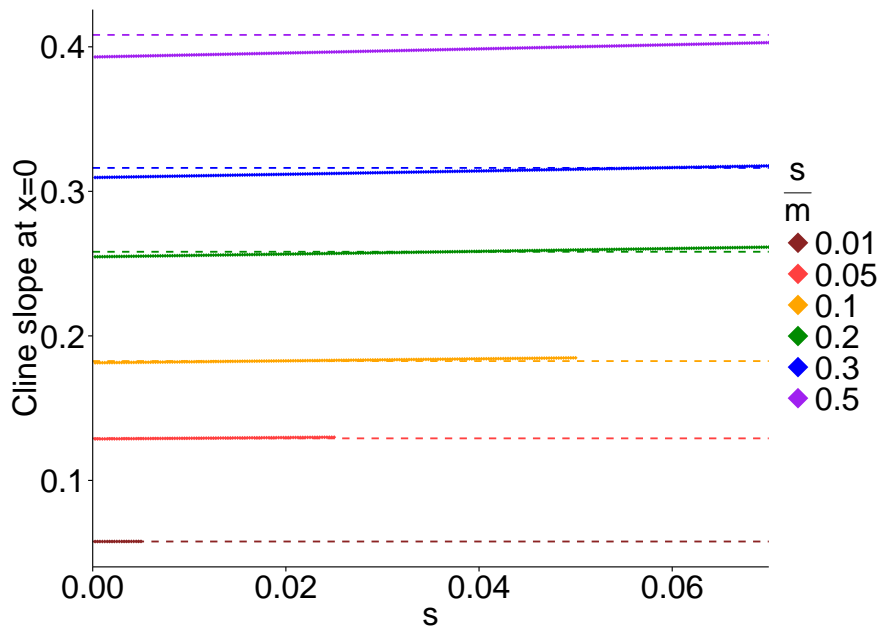
The reason for why deterministic simulations are used when simulation results are compared to the theory in Figure 3.2 is to be able to make a direct comparison between simulation and analytical results (the latter being derived upon neglecting genetic drift). Genetic drift and other stochastic effects shift a cline back and forth around the average midpoint. The effect of genetic drift and how it affects the midpoint of a cline in dependence of selection strength is shown in Figure 3.3. The effect of genetic drift is larger for smaller populations, weaker selection or smaller migration rate (Felsenstein, 1975; Slatkin and Maruyama, 1975) (see Figure 3.3).

The results presented in Figures 3.1 - 3.3 confirmed the validity of the simulations performed in this study. In the next sections, the simulation results obtained under the two models of sexual selection (presented in Chapter 2) are shown.

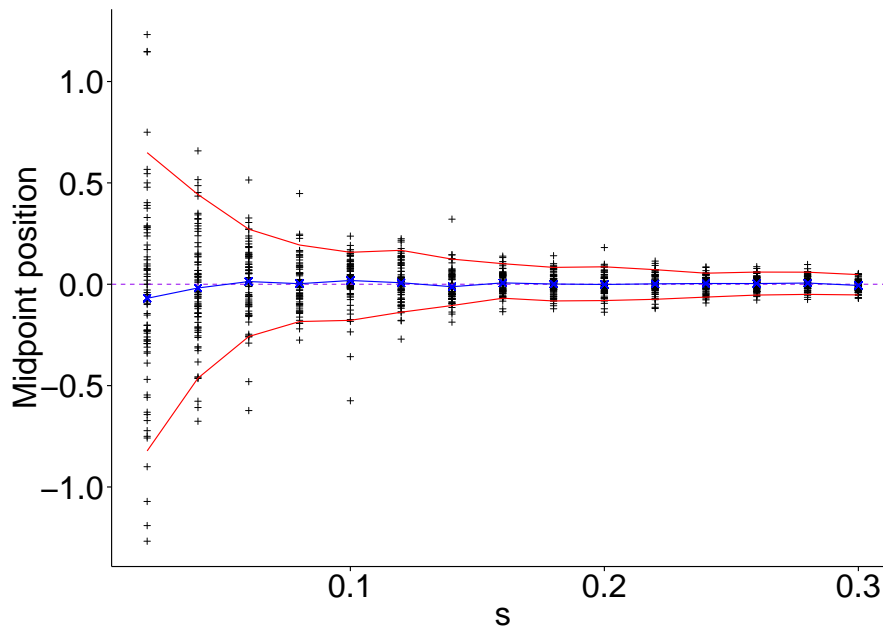




**Figure 3.1:** The frequency of allele  $A$  in each deme (favoured in demes on the left side of the environmental change and disfavoured on the right). These clines are obtained by simulating the model presented by Slatkin (1973), with different values of  $s$  and  $m$ . The clines shown are obtained 5000 generations after the start of the simulations. At the start of each simulation both alleles  $A$  and  $a$  have the same frequency in all demes. The cline at the end of each simulation (grey lines) is plotted together with the average of all these clines (green line) a deterministic simulation of a cline (red line) the theory by Slatkin (1973) (blue line) the tangens hyperbolicus function (pink line) and the expected steady state in the case of no migration (black dashed line). Each panel consists of 500 independent simulations of 50 demes, each with 100 individuals.



**Figure 3.2:** The slope of a cline as a function of selection strength  $s$ . Each simulation is shown as a point and the theoretical slope is shown with a dashed line. Different colours denote different values of  $\frac{s}{m}$  used (see the legend). The effect of genetic drift is ignored in these simulations (the population is assumed to be infinite). The starting condition for all simulations is that both the  $A$  and  $a$  allele have the same frequency in all demes. For a given value of  $s$ , the parameter  $m$  is chosen so that the fraction  $\frac{s}{m}$  is kept constant. The number of demes used in each simulation is 100.



**Figure 3.3:** The effect of genetic drift on the midpoint position of a cline in relation of the environmental change as a function of selection strength,  $s$ . A cross in the plot is the midpoint of a cline obtained 5000 generation after the start of the corresponding simulation. At the start of each simulation the frequency of allele  $A$  and  $a$  is equal in all demes. The average midpoint position over 100 independent simulations (for each parameter set) as a function of  $s$  is shown by a blue line and the red lines in the plot shows the 10% (bottom) and 90% (upper) quantiles. The dashed purple line shows the expected average midpoint position. The reason that the average midpoint position does not coincide with the expected midpoint position for small  $s$  is because the effect of genetic drift is much stronger and thus the variance of the observed samples, the sample size therefore too small to capture the real mean. In all simulations the migration rate is set to 0.1. The population consisted of 50 demes, each with 100 individuals.

## 3.2 The effect of assortative mating

Simulation results obtained under the model of assortative mating, (described in Section 2.3,) show that the midpoint on average coincides with the position of environmental change, (see Figure 3.4) To explain the results shown in Figure 3.4 the gene flow, between the two sides of the change in the environments is considered for two types of assortative mating, as explained next.

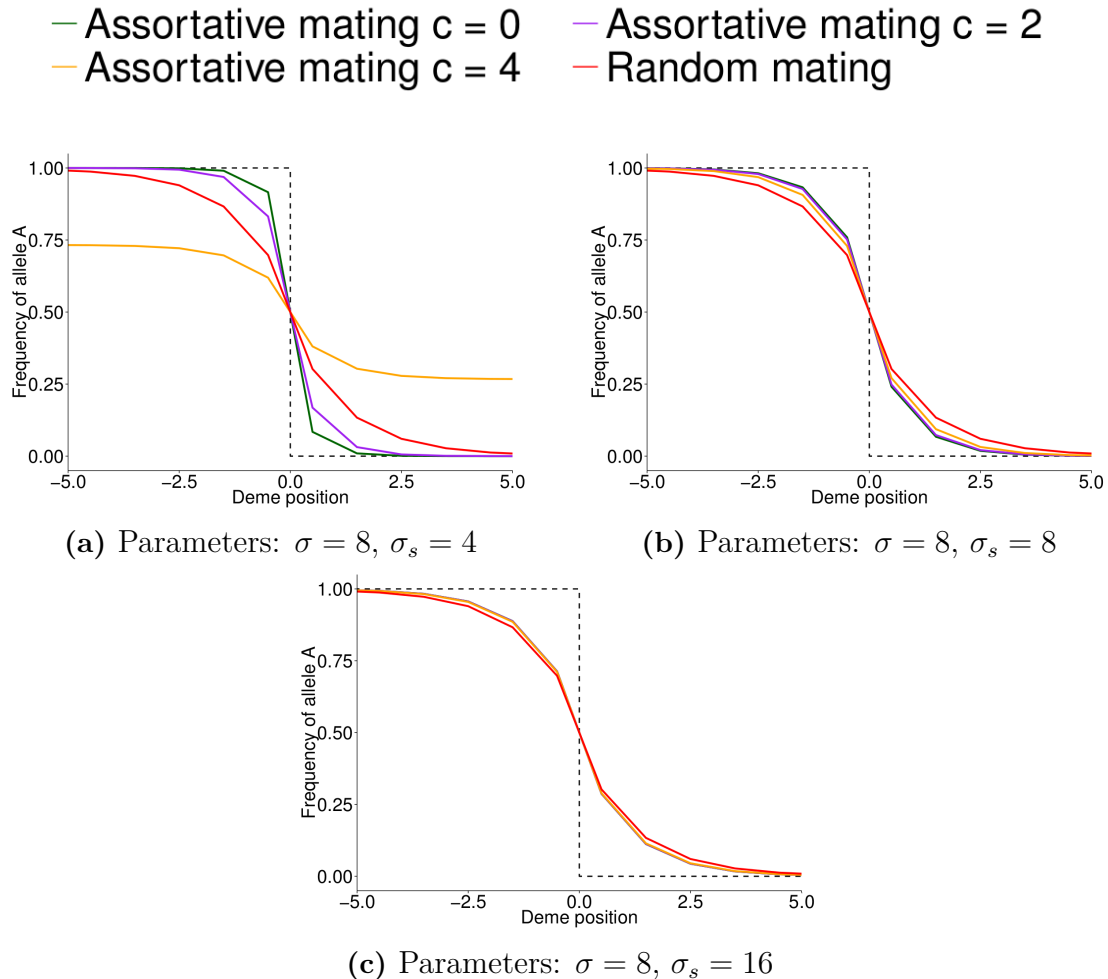
Firstly, in the case where individuals prefer to mate with individuals of the same size as themselves ( $c = 0$ ) the gene flow between the two sides of the environments decreases and heterozygote individuals is rarer. However the gene flow of females and males is the same in both directions. This agrees with Kirkpatrick (2000) who shows that sexual preference for individuals of similar size results in a strengthened barrier to gene flow between the two sides of a cline.

Secondly, if males prefer to mate with females larger than themselves ( $c > 0$ ) the gene flow of females, from the side where larger individuals are favoured to the side where smaller individuals are favoured, is increased. However the gene flow of males from the side where smaller individuals are favoured to the side where larger individuals are favoured, is increased. This occurs since mating preference is mutual (see Equation (2.4)). The gene flow of males and females thus balances each other so that the total gene flow, between the two sides, is symmetric.

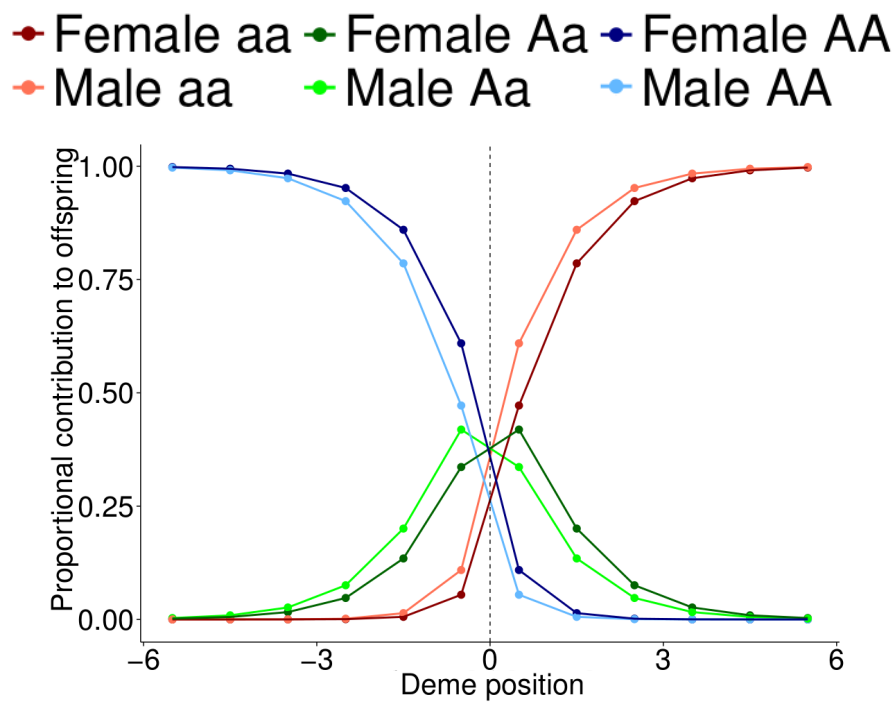
There is no need to consider the third case of  $c < 0$  since similar arguments as in the previous case of where  $c > 0$  apply but with males substituted by females and vice versa.

The difference in the gene flow between males and females can be observed by studying the differences in the distribution of genotypes among males and females that contributes to the next generation (see 3.5).

The results presented in Figure 3.4 and Figure 3.5 show that the assortative mating model presented in this thesis can not produce a shift in the average midpoint position as observed shift in *L. saxatilis*. In the next section the results obtained from the other type of sexual selection (size dependent fecundity of males) considered in this thesis is presented.



**Figure 3.4:** The average frequency of allele  $A$  (favoured in demes on the left side of the environmental change and disfavoured on the right) as a function of deme position for  $\sigma = 8$  and three different values of the strength of attractiveness  $\sigma_s$  (see above). At the start of each simulation the frequency of allele  $A$  and  $a$  is equal in all demes. The clines shown are obtained 5000 generations after the start of the simulation and is the average of 500 independent simulations. Three different mating preferences for assortative mating is presented in the plot:  $c = 0$  (green lines),  $c = 2$  (blue lines),  $c = 4$ , (orange lines). The case where individuals do not have any mating preferences, random mating, is represented by red lines. The expected steady state in the case of no migration and no assortative mating is shown as a black dashed line. The migration rate  $m$  is set to 0.1 in all simulations. There are 80 demes, each with 100 individuals in each simulation. There are no mutations,  $\mu = 0$ .



**Figure 3.5:** The contribution of individual genotypes of males and females, locally within each deme, to the next generation. This figure illustrates the differences in gene flow between males and females as well as the symmetries explained in the text. The results are obtained using deterministic simulation with 50 demes and an infinite number of individuals in each deme. The frequency of allele  $A$  and  $a$  is equal in all demes at the start of the simulation. Parameters used:  $\sigma = 8$ ,  $\sigma_s = 8$  and  $m = 0.1$ ,  $c = 4$ ,  $\mu = 0$ .

### 3.3 The effect of size dependent male fecundity

In this section the simulation results obtained under the model described in Section 2.4 is presented.

The first result is obtained by studying multiple demes with an environmental change in the middle where different alleles are favoured on different sides of the change. Still there are only two alleles in the population, allele  $A$  and  $a$ . Assuming that sexual selection favours males with allele  $A$ , due to an increased fecundity, a stable polymorphism may be obtained on the side of the environmental change where the  $a$  allele is favoured by natural selection (see Appendix A.2 for the conditions under which a stable polymorphism forms). This polymorphism causes a shift in the cline (see Figure 3.6). However if the sexual selection of males, through increased fecundity, is much stronger than the natural selection the allele  $A$  fixates in the whole population and no cline is obtained.

The same model is then simulated with mutations. An example of a typical time evolution of this model is shown in Figure 3.7. The simulation starts from a population where all individuals have a phenotype equal to the average value of the optimal phenotypes set by natural selection in the two different environments. In the transient state of the simulation the average midpoint position of the cline is shifted to the right of the environmental change. The reason a shift occurs during the transient state is that males that are larger have higher fecundity and thus the effective male gene flow from the side of the cline where individuals are larger to the side of the cline where individuals are smaller is increased. However the effective gene flow of females are not affected. Surprisingly, this shift of the average midpoint position decreases over time and when the system reached the steady state no shift in the average midpoint can be observed. The time evolution of the model is shown in Figure 3.8.

A qualitative study is then conducted to find of how each parameter affects the average magnitude of the maximum shift and the number of generations until the steady state is reached. The result is presented in the text below and a summary is presented in Table 3.1. A list of all parameter values that are used in this analysis is shown in Table 3.2. All results are concluded from simulations with the parameter sets from Table 3.2 and thus other effects than the one described below can occur for different parameters.

The first parameter that is studied is  $\sigma$ . An increase in  $\sigma$  decreases the strength of natural selection. When natural selection is weak, the effects of sexual selection is greater. This makes the magnitude of the maximum shift larger. A weaker natural selection also increases the time it takes for an advantageous allele to spread and thus the time until the steady state is reached is increased.

If the average squared dispersal distance  $l$  is increased the gene flow between demes is increased. This increases the maximum magnitude of the shift. For the parameter values tested the time to reach the steady state decreased when the value of  $l$  increased. This is probably due to that an advantageous mutation spreads faster if the average dispersal distance is larger.

When considering how the male-size fecundity parameter  $d$  affects the system two extreme cases is first studied. If  $d = 0$  (that is, no increased fecundity with size), there is no shift of the average midpoint position. If  $d$  is large and the sexual selection is much stronger than the natural selection it is beneficial for individuals in both types of the environment to become ever larger. Therefore no cline forms (and no shift either). Finally, when  $d > 0$ , (but not too large) and natural selection is sufficiently strong a shift in the cline is observed. Thus the maximum magnitude of the shift first increases with an increasing  $d$  and then after some critical  $d$  value the maximum magnitude of the shift decreases with an increasing  $d$ . In the same way as a stronger natural selection decrease the time until the steady state is reached, sexual selection decreases the time until the steady state is reached.

For the parameter values that are tested an increase in the population size  $N$  decreased the average magnitude of the maximum shift. This is probably due to that an increase in the population size decreases the effect of genetic drift. An increase in the population size  $N$  does also decrease the time until the steady state is reached since there is more mutations in each generation the bigger the population is since the mutation rate  $\mu$  is the mutation rate per allele. Changing the mutation rate  $\mu$  or the standard deviation of the average mutation effect size  $\mu_{sd}$  does not effect the maximum magnitude of the shift or the average allele-effect size at the steady state.

Finally, note that due to sexual selection the phenotype values obtained at the the steady state are larger than the optimal phenotype set by natural selection. As a consequence of this effect, the shift observed during a transient state disappears when the system reaches the steady state. This is further explained in Chapter 4.

In this chapter the results from simulations of the models described in Chapter 2 are presented. To conclude this chapter, it is found that the first model with assortative mating does not produce a shift. However the second model where male fecundity depends on size produced a shift during the transient state. The results from both these models are further discussed and interpreted in the next chapter.



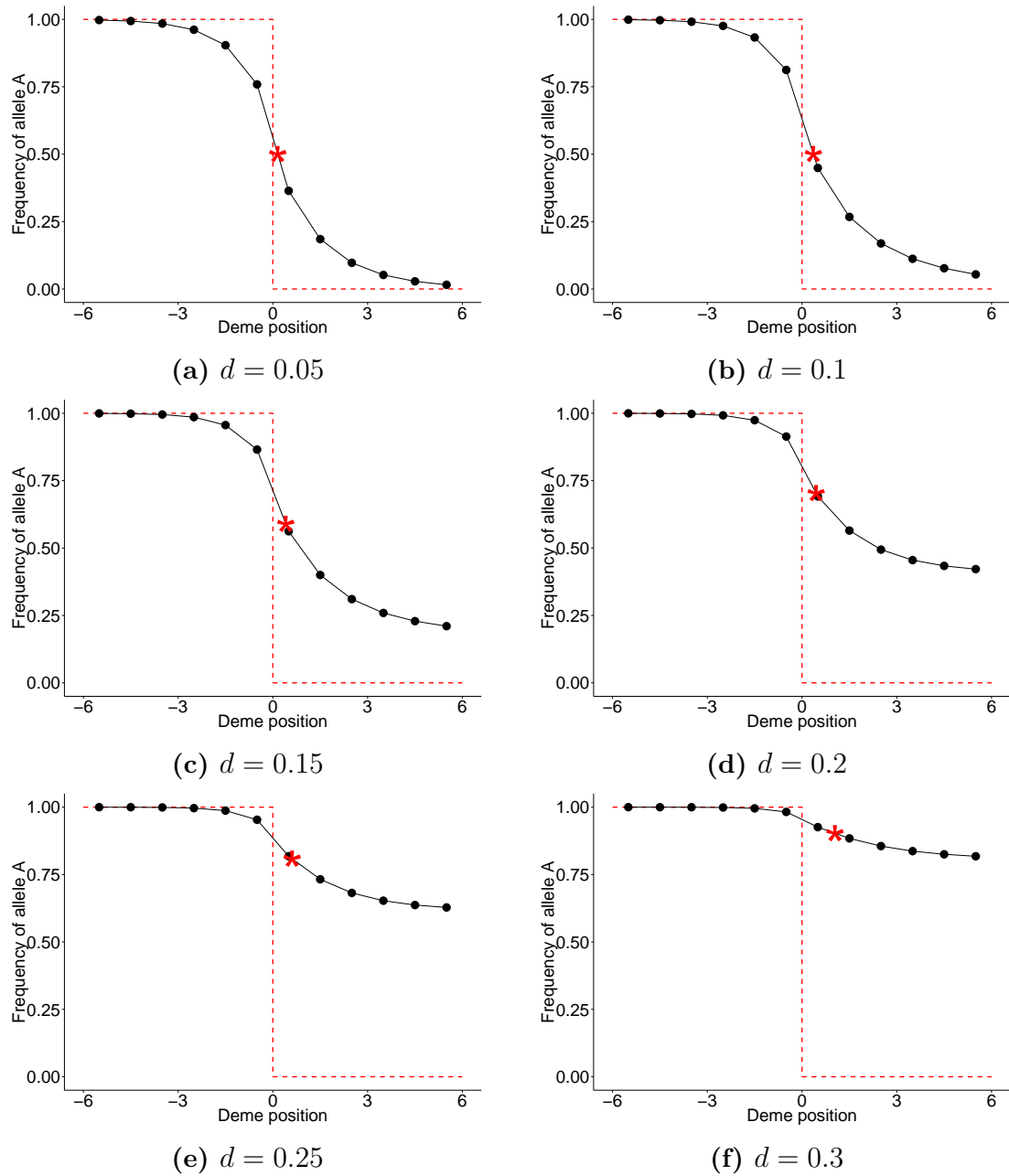
Increase in:	Effect on the maximum magnitude of the shift	Effect of time until the steady state is reached
$\sigma$	Increase	Increases
$l$	Increase	Decreases
$d$	First increases then decreases	Decreases
$N$	Decrease	Decreases
$\mu$	Unchanged	Decreases
$\mu_{sd}$	Unchanged	Decrease

**Table 3.1:** The qualitative change of the system then a given parameter is increased. See text above for further description.

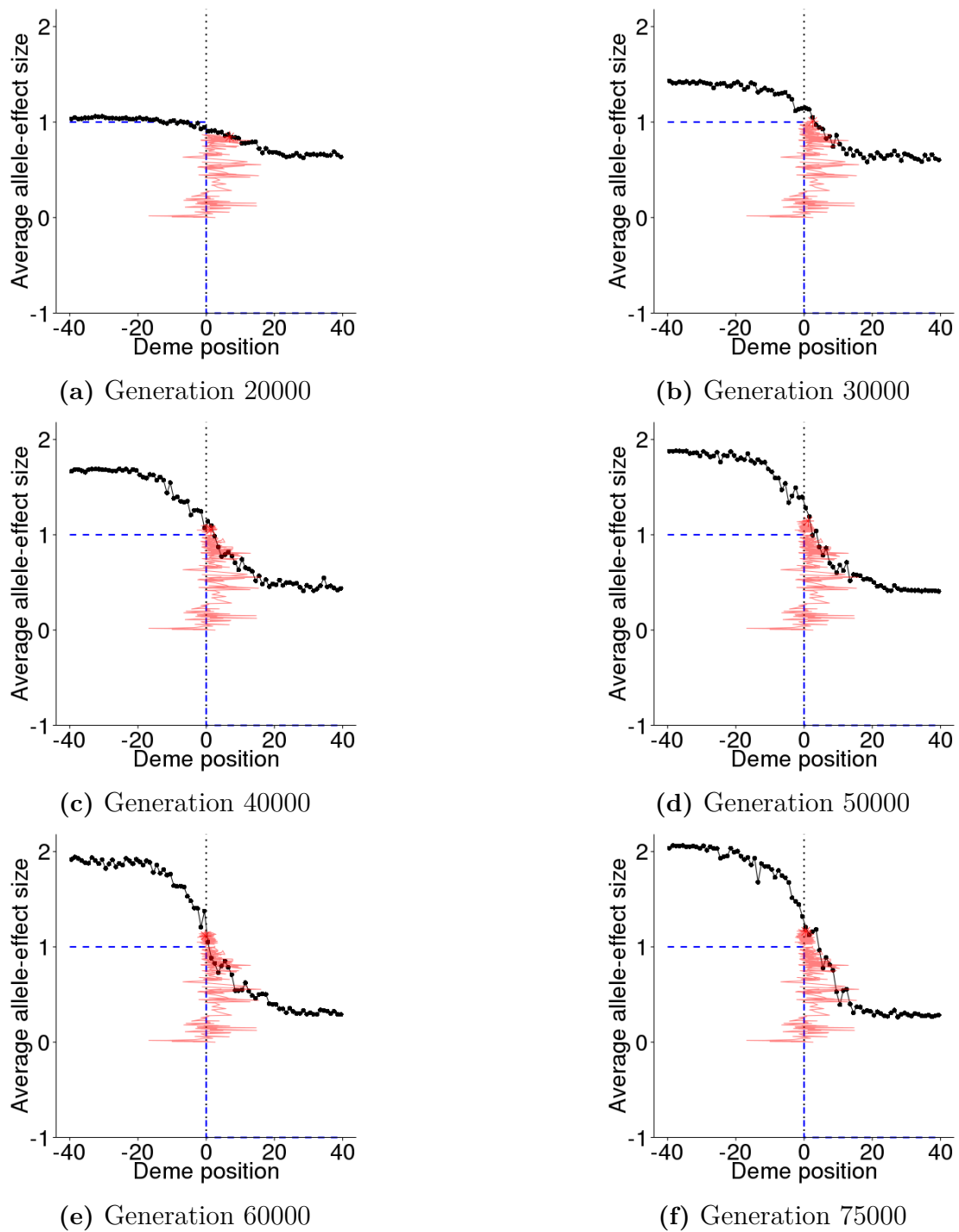
Parameter	Values
$\sigma$	6, <b>8</b> , 12, 16
$l$	0.5, <b>1</b> , 2
$d$	0, <b>0.1</b> , 0.25, 0.5, 0.8
$N$	50, <b>100</b> , 200
$\mu$	<b>0.0002</b> , 0.0004, 0.001
$\mu_{sd}$	0.01, <b>0.05</b> , 0.1

**Table 3.2:** Parameters used to determine the qualitative behaviour of the model. Each parameter is varied one at a time. Values shown in bold font is the default value that is used when other parameters are varied.

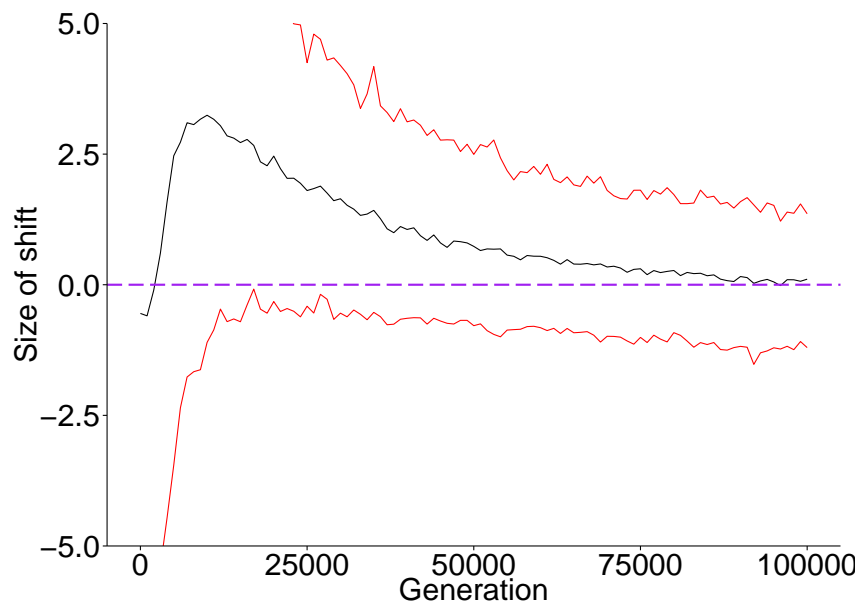
### 3. Results



**Figure 3.6:** The average frequency of allele  $A$  (favoured in demes on the left side of the environmental change and disfavoured on the right) as a function of deme position. Results from simulations are presented with black points and the midpoint of the cline is shown as a red star. The effect of genetic drift is ignored in these simulations since the population is assumed to be infinite and these simulations are thus deterministic. The starting condition for all simulations is that both the  $A$  and  $a$  allele have the same frequency in all demes. The expected steady state in the case of no migration and size independent fecundity ( $d = 0$ ) is shown with a red dashed line. There are 50 demes in each simulation. The population size  $N$  is infinitely large and the mutation rate  $\mu = 0$ . In all panels the natural selection parameter  $\sigma = 8$ , and the migration rate  $m = 0.1$ . The  $d$  value is varied between 0.05 and 0.3, the  $d$  value used is presented under each panel.



**Figure 3.7:** Average allele-effect size as a function of deme position under the model in which fecundity of a male is larger when the male is larger. This is a typical time evolution of a population. The red line is the trace of midpoints through time and the midpoint at the current generation is represented with a red star. The expected steady state in the case of no migration and size independent fecundity ( $d = 0$ ) is shown as a blue dashed line. The black dotted line show the centre of the population where the environmental shift occurs. Parameters:  $\sigma = 8$ ,  $l = 2$  and  $d = 0.1, \mu = 0.0002$ ,  $\mu_{sd} = 0.05$ ,  $M = 80$  and  $N = 100$ .



**Figure 3.8:** The shift in the midpoint of the average cline as a function of time. Each point is the average over 25 independent simulations. The purple dashed line represents no shift at all from the environmental change in the habitat. The red lines in the plot show the 10% (bottom) and 90% quantiles (top). Parameters:  $\sigma = 8$ ,  $l = 2$ ,  $d = 0.1$ ,  $\mu = 0.0002$ ,  $\mu_{sd} = 0.05$ ,  $M = 80$  and  $N = 100$ .

# 4

## Discussion and Conclusions

Different forms of sexual selection are present in biological populations. An example is *L. saxatilis* where both assortative mating and size-dependent fecundity of males are observed (Johannesson et al., 1995; Erlandsson and Rolán-Alvarez, 1998; Hintz Saltin et al., in press; Johannesson and Butlin, personal communication). This thesis aims to investigate the effects these types of sexual selection have on clines. To this end, two models are created to resemble the empirical data presented by Johannesson and Butlin (personal communication) and Hintz Saltin et al. (in press).

The first results presented in this thesis are obtained by simulating the model analysed by Slatkin (1973) but with a finite population size. Note that the two models presented in this thesis reduce to the model by Slatkin (1973) in following limits: selection is weak ( $\sigma$  is large), no mutations occur ( $\mu = 0$ ), no assortative mating ( $\sigma_s \rightarrow \infty$ ), and fecundity of males is independent of size ( $d = 0$ ). To validate the simulated results and evaluate in which limits the theory is valid, the slope of simulated clines is compared to the theoretical slope derived by Slatkin (1973). In this derivation by Slatkin (1973) it is assumed that both selection and migration are weak and that the local population size is infinitely large ( $N \rightarrow \infty$ ) as well as that the population is continuously distributed. As expected this thesis finds that when the population is divided into discrete demes and ordered in a one dimensional array the theory for the midpoint of the cline is valid, provided that the limits mentioned above are fulfilled, as well as the selection is sufficient weaker than migration (but not too weak).

The model by Slatkin (1973) is then extended to a model with assortative mating. However no shift in the average midpoint position of the cline is observed in this model. This is because the model is perfectly symmetric since the preference of females among males is exactly the opposite of the preference of males among females.

In addition to assortative mating there is another type of sexual selection present in *L. saxatilis*: size-dependent male fecundity. The use of this model is motivated by recent findings by Hintz Saltin et al. (in press), that are confirmed in this thesis (see Appendix B). In this model a shift of the midpoint of a cline occurs during a transient state, which lasts for up to  $10^5$  generations. This is a longer time period than the time that has past since the end of the last glacial period, when *L. saxatilis* started to spread in Sweden. Thus it is likely that *L. saxatilis* is in a transient state and the shift of the midpoint of a cline may be significant according to the model

results shown in this thesis.

Indeed, in the transient state the effective selection for larger individuals is effectively stronger in the environment where both natural and sexual selection favours large individuals, than selection for smaller individuals in the opposite part of the habitat. Thus the model results are effectively captured by a model with asymmetric selection pressure in the two parts of the habitat, such as the model by Haldane (1948). However, the model studied here allows for understanding the temporal dynamics of the interplay between natural and sexual selection, resulting in an intriguing evolution of cline patterns over time. Indeed, as this thesis shows the midpoint of the cline is shifted from the geographical position of the environmental change but only during a transient state. The shift grows initially in its magnitude, reaching a maximum value, and thereafter it progressively decays towards zero. The reason for why the shift decays towards zero is the following. Recall that, due to sexual selection favouring large individuals, the phenotype values obtained at the steady state are larger than the optimal phenotype values set by natural selection. Therefore natural selection against small individuals in the environment, where large individuals are favoured becomes weaker than natural selection against large individuals in the opposite environment. However, sexual selection increases the effective selection pressure against small individuals in their disfavoured environment, and decreases the effective selection pressure against large individuals in their disfavoured environment. This joint effect of natural and sexual selection of the system, gives rise to effectively similar selection pressures in the two environments in the steady state. As a result, effective gene flow in the steady state becomes symmetric between the opposite environments, and so the shift of the position of the average midpoint of a cline from the geographic position of the environmental change becomes negligible.

The models presented in this thesis can be modified further to account for more complex and biologically relevant scenarios. For example, the model where male fecundity depends on size that is analysed here, does not address the evolution of clines in other traits that are under natural selection. To address this question it is necessary to add more traits to the model. To this end, two extreme cases can be considered. The first case is to assume that the additional trait is completely unlinked to the size trait. In this case the cline in this additional trait and the cline in size are completely independent and thus the cline in size is expected to be shifted in relation to this new trait. The shifts of the two clines will therefore be different. The second extreme case is when this additional trait is completely linked to size trait. In this case the cline in size and the cline in this additional trait are fully linked and no difference between them is expected (see Kruuk et al. (1999) and Hare and Avise (1996) for models with multiple traits but without sexual selection). It still remains to be explained what happens for arbitrary levels of genetic linkage (between the two extreme cases mentioned above). It is especially important to find at what level of linkage there will be a significant difference in the position of the clines in the different traits. This will allow to make a direct comparison to the corresponding linkage estimates from the soon-to-come empirical data, thanks to the ongoing genetic studies within the Centre for Marine Evolutionary Biology, University of Gothenburg. Finally, in the models analysed here it is assumed that

the strength of selection and the optimal phenotypes are constant during time. This is, however, not the case in nature where the environment is changing over time. The effect of changes in environmental conditions on the evolution of clines remains to be understood in further studies.

In conclusion, the model analysed in the thesis seems a good candidate to explain the empirically observed shifts in the size clines in *L. saxatilis*, but as discussed in this chapter the finding needs to be further supported by additional analyses of clines in other traits tightly or loosely linked to the size trait. This remains to be analysed in a future work.





# Bibliography

- Barton, N. H. and Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual review of Ecology and Systematics*, 16:113–148.
- Barton, N. H. and Hewitt, G. M. (1989). Adaptation, speciation and hybrid zones. *Nature*, 341(6242):497–503.
- Bazykin, A. (1969). Hypothetical mechanism of speciation. *Evolution*, 23:685–687.
- Erlandsson, J. and Rolán-Alvarez, E. (1998). Sexual selection and assortative mating by size and their roles in the maintenance of a polymorphism in Swedish *Littorina saxatilis* populations. *Hydrobiologia*, 378(1-3):59–69.
- Felsenstein, J. (1975). Genetic drift in clines which are maintained by migration and natural selection. *Genetics*, 81(1):191–207.
- Felsenstein, J. (1976). The theoretical population genetics of variable selection and migration. *Annual review of genetics*, 10(1):253–280.
- Fife, P. C. (1979). *Mathematical aspects of reacting and diffusing systems*, volume 28. Springer-Verlag, New York;Berlin;.
- Fisher, R. A. (1937). The wave of advance of advantageous genes. *Annals of Eugenics*, 7(4):355–369.
- Fitzpatrick, B. M. (2013). Alternative forms for genomic clines. *Ecology and evolution*, 3(7):1951–1966.
- Galindo, H. M., Pfeiffer-Herbert, A. S., McManus, M. A., Chao, Y., Chai, F., and Palumbi, S. R. (2010). Seascape genetics along a steep cline: using genetic patterns to test predictions of marine larval dispersal. *Molecular Ecology*, 19(17):3692–3707.
- Haldane, J. (1948). The theory of a cline. *Journal of genetics*, 48(3):277–284.
- Hare, M. P. and Avise, J. C. (1996). Molecular genetic analysis of a stepped multilocus cline in the american oyster (*Crassostrea virginica*). *Evolution*, 50:2305–2315.
- Hintz Saltin, S., Charrier, G., Panova, M., Ring, A.-K., Kvarnemo, C., André, C., and Johannesson, K. (In press). Sexual selection in a highly outcrossing marine snail.

- Janson, K. (1983). Selection and migration in two distinct phenotypes of *Littorina saxatilis* in Sweden. *Oecologia*, 59(1):58–61.
- Johannesson, K., Rolan-Alvarez, E., and Ekendahl, A. (1995). Incipient reproductive isolation between two sympatric morphs of the intertidal snail *Littorina saxatilis*. *Evolution*, 14:1180–1190.
- Johannesson, K. and Warmoes, T. (1990). Rapid colonization of belgian breakwaters by the direct developer, *Littorina saxatilis* (olivi)(prosobranchia, mollusca). In *Progress in Littorinid and Muricid Biology*, volume 193, : 99–108. Springer.
- Johnson, N. (2008). Hybrid incompatibility and speciation. *Nature Education*, 1(1):20.
- Kimura, K., Hirano, T., and Chiba, S. (2015). Assortative mating with respect to size in the simultaneously hermaphroditic land snail *bradybaena pellucida*. *acta ethologica*, 18(3):265–268.
- Kirkpatrick, M. (2000). Reinforcement and divergence under assortative mating. *Proceedings of the Royal Society of London B: Biological Sciences*, 267(1453):1649–1655.
- Kruuk, L., Baird, S., Gale, K., and Barton, N. (1999). A comparison of multilocus clines maintained by environmental adaptation or by selection against hybrids. *Genetics*, 153(4):1959–1971.
- Lande, R. (1981). Models of speciation by sexual selection on polygenic traits. *Proceedings of the National Academy of Sciences*, 78(6):3721–3725.
- Mallet, J., Barton, N., Lamas, G., Santisteban, J., Muedas, M., and Eeley, H. (1990). Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in *heliconius* hybrid zones. *Genetics*, 124(4):921–936.
- May, R. M., Endler, J. A., and McMurtrie, R. E. (1975). Gene frequency clines in the presence of selection opposed by gene flow. *American Naturalist*, 109:659–676.
- Nagylaki, T. (1978). Clines with asymmetric migration. *Genetics*, 88(4):813–827.
- Polechová, J. and Barton, N. (2011). Genetic drift widens the expected cline but narrows the expected cline width. *Genetics*, 189(1):227–235.
- Sadedin, S., Hollander, J., Panova, M., Johannesson, K., and Gavrillets, S. (2009). Case studies and mathematical models of ecological speciation. 3: Ecotype formation in a Swedish snail. *Molecular Ecology*, 18(19):4006–4023.
- Slatkin, M. (1973). Gene flow and selection in a cline. *Genetics*, 75(4):733–756.
- Slatkin, M. and Maruyama, T. (1975). Genetic drift in a cline. *Genetics*, 81(1):209–222.

# A

## Mathematical derivations

### A.1 Slope of a cline

In this section the theoretical slope of a cline for the model described by Slatkin (1973) is derived. This work is derived following in a similar way as Haldane (1948) derived the slope of a cline but with the model presented by Slatkin (1973). First of all the model that is analysed in this section and is presented by Slatkin (1973) is described. The assumptions that are made in the model are:

- The number of individuals, denoted  $N$ , is assumed to be equal in all demes at the start of each generation.
- Individuals are diploid and assumed to have one locus under selection. This locus have two possible alleles ( $A$  and  $a$ ) so each individual have one of the genotypes  $AA$ ,  $Aa$  or  $aa$ .
- The environmental conditions vary in the habitat. At the midpoint of the habitat there is an abrupt change of the environmental conditions. On the left side of this change the allele  $A$  is beneficial and on the right side the allele  $a$  is beneficial for individuals to have.
- Generations are assumed to be discrete and non-overlapping.
- In each generation, it is assumed that the following process occur:
  1. maturation and migration,
  2. selection,
  3. death and beginning of the next generation.
- Locally within each deme the number of surviving offspring to maternity is assumed to be  $N$ . The chance to survive to maternity is assumed to be independent of genotype and location of an individual.
- Individuals may migrate from the habitat they are born in. The squared average dispersal distance is denoted  $l$  and the dispersal is symmetric in both directions.

- Natural selection is assumed to occur after migration. The natural selection considered is viability selection, where the probability for an individual to survive to mate is proportional to its fitness. The fitness of the different genotypes at deme number  $k$  is given in Table A.1.

Deme position \ Genotype	AA	Aa	aa
$k < \frac{M+1}{2}$	$1 + s$	1	$1 - s$
$k > \frac{M+1}{2}$	$1 - s$	1	$1 + s$

**Table A.1:** The selective pressure for the three genotypes in the different parts of the population. Here  $s$  is the selection parameter that determines the strength of selection.

Furthermore, in the mathematical analyses of the system the two additional assumptions are made:

- The population lives in an infinite two dimensional space and the population density is equal throughout this space.
- The local population size is infinitely large,  $N \rightarrow \infty$ .

The frequency of allele  $A$  at position  $x$  at time  $t$  is denoted by  $p(x, t)$  and  $Q_x(\delta x)$  is the probability that an individual migrates to the position  $x$  from  $x + \delta x$  in time  $\delta t$ . It is assumed that migration is not directional so  $Q_x$  is symmetric around 0. Only considering migration the frequency of allele  $A$  at time  $t$  at position  $x$  can then be described by the equation:

$$p(x, t) = \int_{-\infty}^{\infty} Q_x(\delta x) p(x - \delta x, t - \delta t) d\delta x. \quad (\text{A.1})$$

Expanding  $p(x - \delta x, t - \delta t)$  in a Taylor series gives

$$p(x - \delta x, t - \delta t) = p(x, t) - \delta x \frac{\partial p}{\partial x} - \delta t \frac{\partial p}{\partial t} + \frac{\delta x^2}{2} \frac{\partial^2 p}{\partial x^2} + \delta x \delta t \frac{\partial^2 p}{\partial x \partial t} + \frac{\delta t^2}{2} \frac{\partial^2 p}{\partial t^2} + \dots \quad (\text{A.2})$$

Inserting Equation (A.2) in Equation (A.1) and using the assumption that  $Q_x$  is symmetric gives

$$p(x, t) = p(x, t) - \delta t \frac{\partial p}{\partial t} + \frac{\delta t^2}{2} \frac{\partial^2 p}{\partial t^2} + \frac{\partial^2 p}{\partial x^2} \int_{-\infty}^{\infty} \frac{\delta x^2}{2} Q_x(\delta x) d\delta x + \dots \quad (\text{A.3})$$

It is now assumed that  $Q_x$  is a Gaussian distributed and defined as

$$Q_x(\delta x) = \frac{1}{\sqrt{l} \sqrt{2\pi}} e^{-\frac{\delta x^2}{2l}}, \quad (\text{A.4})$$

where sigma is the standard deviation. Using this definition Equation (A.3) is rewritten to

$$\begin{aligned} p(x, t) &= p(x, t) - \delta t \frac{\partial p}{\partial t} + \frac{\delta t^2}{2} \frac{\partial^2 p}{\partial t^2} + \frac{\partial^2 p}{\partial x^2} \int_{-\infty}^{\infty} \frac{\delta x^2}{2} \frac{1}{\sqrt{l}\sqrt{2\pi}} e^{-\frac{\delta x^2}{2l}} d\delta x + \dots \\ &= p(x, t) - \delta t \frac{\partial p}{\partial t} + \frac{\delta t^2}{2} \frac{\partial^2 p}{\partial t^2} + \frac{l}{2} \frac{\partial^2 p}{\partial x^2} + \dots \end{aligned} \quad (\text{A.5})$$

It is now time to focus on how selection affects the frequency of allele  $A$ . For  $x > 0$  the change in frequency in time  $\delta t$  can be written as

$$p(x, t) = \frac{sp(x, t - \delta t)^2 + p(x, t - \delta t)}{1 + s(2p(x, t - \delta t) - 1)}. \quad (\text{A.6})$$

Assuming  $s$  is small Equation (A.6) is approximated as

$$p(x, t) \approx p(x, t - \delta t) + s(p(x, t - \delta t) - p(x, t - \delta t)^2). \quad (\text{A.7})$$

The same is done for  $x < 0$  giving the following approximation

$$p(x, t) \approx p(x, t - \delta t) - s(p(x, t - \delta t) - p(x, t - \delta t)^2). \quad (\text{A.8})$$

Inserting the change in frequency of allele  $A$  due to selection as described by Equation (A.7) in Equation (A.5) gives:

$$\begin{aligned} p(x, t) &= p(x, t - \delta t) + s(p(x, t - \delta t) - p(x, t - \delta t)^2) \\ &\quad - \delta t \frac{\partial p}{\partial t} + \frac{\delta t^2}{2} \frac{\partial^2 p}{\partial t^2} + \frac{l}{2} \frac{\partial^2 p}{\partial x^2} + \dots \end{aligned} \quad (\text{A.9})$$

It is now assumed that the system is in an equilibrium where the frequency of allele  $A$  is kept constant over time and thus  $p(x, t) = p(x, t - \delta t)$ . Since the frequency of allele  $A$  is kept constant all derivatives over time would be 0. With this assumption Equation (A.9) can be rewritten as

$$p(x, t) = p(x, t) + s(p(x, t) - p(x, t)^2) + \frac{l}{2} \frac{d^2 p}{dx^2}. \quad (\text{A.10})$$

The same approach is done for  $x < 0$  and inserting Equation (A.8) in Equation (A.5) gives:

$$p(x, t) = p(x, t) - s(p(x, t) - p(x, t)^2) + \frac{l}{2} \frac{d^2 p}{dx^2}. \quad (\text{A.11})$$

Reordering Equation (A.10) and (A.11) gives the two differential equations:

$$\frac{\partial^2 p}{\partial x^2} = -\frac{2}{l}s(p(x, t) - p(x, t)^2), \quad (\text{A.12})$$

$$\frac{\partial^2 p}{\partial x^2} = \frac{2}{l}s(p(x, t) - p(x, t)^2). \quad (\text{A.13})$$

To solve these equations the variable  $q = \frac{dp}{dx}$  is introduced. Equation (A.12) is then rewritten as:

$$q \frac{dq}{dp} = -\frac{2}{l}s(p(x, t) - p(x, t)^2), \quad (\text{A.14})$$

$$\int q dq = -\int \frac{2}{l}s(p(x, t) - p(x, t)^2) dp, \quad (\text{A.15})$$

$$\frac{1}{2}q^2 = -\frac{2}{l}s\left(\frac{p(x, t)^2}{2} - \frac{p(x, t)^3}{3}\right) + C, \quad (\text{A.16})$$

$$q^2 = -\frac{4}{l}s\left(\frac{p(x, t)^2}{2} - \frac{p(x, t)^3}{3}\right) + 2C. \quad (\text{A.17})$$

The border condition that when  $x \rightarrow \infty$   $p(x, t) \rightarrow 1$  and  $q \rightarrow 0$  is then used. This gives that  $C = \frac{s}{3l}$  and

$$q^2 = -\frac{4}{\sigma^2}s\left(\frac{p(x, t)^2}{2} - \frac{p(x, t)^3}{3} - \frac{1}{6}\right), \quad (\text{A.18})$$

$$\frac{dp}{dx} = \sqrt{-\frac{4}{l}s\left(\frac{p(x, t)^2}{2} - \frac{p(x, t)^3}{3} - \frac{1}{6}\right)}. \quad (\text{A.19})$$

The same is done for Equation (A.13) that is valid for  $x < 0$  giving the equation

$$\frac{dp}{dx} = \sqrt{\frac{4}{l}s\left(\frac{p(x, t)^2}{2} - \frac{p(x, t)^3}{3}\right)}. \quad (\text{A.20})$$

At  $x = 0$  (the midpoint of the cline) Equation (A.19) and Equation (A.20) must have the same value so

$$-\frac{4}{l}s\left(\frac{p(0, t)^2}{2} - \frac{p(0, t)^3}{3} - \frac{1}{6}\right) = \frac{4}{l}s\left(\frac{p(0, t)^2}{2} - \frac{p(0, t)^3}{3}\right), \quad (\text{A.21})$$

$$\frac{1}{6} = \sqrt{2\left(\frac{p(0, t)^2}{2} - \frac{p(0, t)^3}{3}\right)}. \quad (\text{A.22})$$

Equation A.22 has exactly one solution where  $p(0, t)$  is between 0 and 1 and that is  $p(0, t) = \frac{1}{2}$ . Using this value in Equation (A.19) the slope of the midpoint of the cline becomes

$$\begin{aligned} \frac{dp}{dx} &= \sqrt{-\frac{4}{l}s \left( \frac{(1/2)^2}{2} - \frac{(1/2)^3}{3} - \frac{1}{6} \right)}, \\ &= \sqrt{\frac{s}{l} \left( \frac{2}{3} + \frac{1}{6} - \frac{1}{2} \right)}, \\ &= \sqrt{\frac{s}{3l}}. \end{aligned} \tag{A.23}$$

As expected this is the same theoretical slope as found by Slatkin (1973). A comparison between the result in from Equation (A.23) and simulated results are shown and discussed in the main text.

## A.2 Analyse of the polymorphic steady state

### A.2.1 Conditions for the existence of a polymorphic steady state

In this section the conditions for the existence of a polymorphism in the habitat of small individuals in the model where male fecundity depends on size (see Section 2.4) are analysed. Here it is also assumed that the population in the model consists of one single well mixed population.

Let the frequency of gametes that carry genotype  $A$  and are produced in generation  $t$  be denoted  $\gamma_m(t)$  for males and  $\gamma_f(t)$  for females. The frequency of the different genotypes in male and females in generation  $t + 1$  is written as

$$\begin{aligned} p_m(AA, t + 1) &= \gamma_m(t) * \gamma_f(t), \\ p_m(Aa, t + 1) &= (\gamma_m(t) * (1 - \gamma_f(t)) + (1 - \gamma_m(t)) * \gamma_f(t)), \\ p_m(aa, t + 1) &= (1 - \gamma_m(t)) * (1 - \gamma_f(t)), \\ p_f(AA, t + 1) &= \gamma_m(t) * \gamma_f(t), \\ p_f(Aa, t + 1) &= (\gamma_m(t) * (1 - \gamma_f(t)) + (1 - \gamma_m(t)) * \gamma_f(t)), \\ p_f(aa, t + 1) &= (1 - \gamma_m(t)) * (1 - \gamma_f(t)). \end{aligned} \tag{A.24}$$

From Equation (A.24) it is possible to see that the frequency of genotypes in the next generation are the same in both males and females.

To have a polymorphic steady state selection and the increased fecundity of males must work in opposing directions. From this observation the fitness  $\omega_z$  of phenotype  $z$  is assumed to be:

## A. Mathematical derivations

---

$$\frac{\omega_{AA}}{1-s} \mid \frac{\omega_{Aa}}{1} \mid \frac{\omega_{aa}}{1+s},$$

and the fecundity  $\omega_z^{(A)}$  of males with phenotype  $z$  is assumed to be:

$$\frac{\omega_{AA}^{(A)}}{1} \mid \frac{\omega_{Aa}^{(A)}}{1-\frac{d}{2}} \mid \frac{\omega_{aa}^{(A)}}{1-d}.$$

In every generation individuals produce gametes that form the next generation. The number of gametes an individual creates depends on the fitness of that individual. The number of gametes males produce do also depend on their fecundity. The frequency of male gametes with genotype  $A$  in generation  $t+1$  is written as

$$\gamma_m(t+1) = \frac{\omega_{AA}\omega_{AA}^{(A)}p_m(AA, t) + \omega_{Aa}\omega_{Aa}^{(A)}\frac{1}{2}p_m(Aa, t)}{\omega_{AA}\omega_{AA}^{(A)}p_m(AA, t) + \omega_{Aa}\omega_{Aa}^{(A)}p_m(Aa, t) + \omega_{aa}\omega_{aa}^{(A)}p_m(aa, t)}. \quad (\text{A.25})$$

The frequency of female gametes with genotype  $A$  in generation  $t+1$  is written as

$$\gamma_f(t+1) = \frac{\omega_{AA}p_f(AA, t) + \omega_{Aa}p_f(Aa, t)}{\omega_{AA}p_f(AA, t) + \omega_{Aa}p_f(Aa, t) + \omega_{aa}p_f(aa, t)}. \quad (\text{A.26})$$

Combining Equation (A.24) with Equation (A.25) and Equation (A.26) gives the recursive equations for how the frequency of male and female gametes with genotype  $A$  changes over time:

$$\gamma_f(t+1) = \frac{\omega_{AA}\gamma_m\gamma_f + \omega_{Aa}\frac{1}{2}(\gamma_m(1-\gamma_f) + (1-\gamma_m)\gamma_f)}{\omega_{AA}\gamma_m\gamma_f + \omega_{Aa}(\gamma_m(1-\gamma_f) + (1-\gamma_m)\gamma_f) + \omega_{aa}(1-\gamma_m)(1-\gamma_f)}, \quad (\text{A.27})$$

$$\gamma_m(t+1) = \frac{\omega_{AA}\omega_{AA}^{(A)}\gamma_m\gamma_f + \omega_{Aa}\omega_{Aa}^{(A)}\frac{1}{2}(\gamma_m(1-\gamma_f) + (1-\gamma_m)\gamma_f)}{\omega_{AA}\omega_{AA}^{(A)}\gamma_m\gamma_f + \omega_{Aa}\omega_{Aa}^{(A)}(\gamma_m(1-\gamma_f) + (1-\gamma_m)\gamma_f) + \omega_{aa}\omega_{aa}^{(A)}(1-\gamma_m)(1-\gamma_f)}. \quad (\text{A.28})$$

The next step is to find the steady state where  $\gamma_m^*(t+1) = \gamma_m^*(t)$  and  $\gamma_f^*(t+1) = \gamma_f^*(t)$ . Assuming the system is in a steady state equation A.27 can be written as:

$$\frac{\omega_{AA}\gamma_m^*\gamma_f^* + \omega_{Aa}\frac{1}{2}(\gamma_m^*(1-\gamma_f^*) + (1-\gamma_m^*)\gamma_f^*)}{\omega_{AA}\gamma_m^*\gamma_f^* + \omega_{Aa}(\gamma_m^*(1-\gamma_f^*) + (1-\gamma_m^*)\gamma_f^*) + \omega_{aa}(1-\gamma_m^*)(1-\gamma_f^*)} - \gamma_f^* = 0. \quad (\text{A.29})$$



Inserting values for  $\omega_z$  gives

$$\frac{(1-s)\gamma_m^*\gamma_f^* + \frac{1}{2}(\gamma_m^*(1-\gamma_f^*) + (1-\gamma_m^*)\gamma_f^*)}{(1-s)\gamma_m^*\gamma_f^* + (\gamma_m^*(1-\gamma_f^*) + (1-\gamma_m^*)\gamma_f^*) + (1+s)(1-\gamma_m^*)(1-\gamma_f^*)} - \gamma_f^* = 0. \quad (\text{A.30})$$

Reordering equation A.30 gives

$$\gamma_m^* = \gamma_f^* (1 + 2s - 2\gamma_f^*). \quad (\text{A.31})$$

The expression for  $\gamma_m^*$  from equation A.31 is inserted in equation A.28 giving

$$\begin{aligned} & \left( (1-s) (\gamma_f^* (1 + 2s - 2\gamma_f^*)) \gamma_f^* \right. \\ & \quad \left. + (1 - \frac{d}{2}) \frac{1}{2} \left( (\gamma_f^* (1 + 2s - 2\gamma_f^*)) (1 - \gamma_f^*) + (1 - (\gamma_f^* (1 + 2s - 2\gamma_f^*))) \gamma_f^* \right) \right) / \\ & \left( (1-s) (\gamma_f^* (1 + 2s - 2\gamma_f^*)) \gamma_f^* \right. \\ & \quad \left. + (1 - \frac{d}{2}) \left( (\gamma_f^* (1 + 2s - 2\gamma_f^*)) (1 - \gamma_f^*) + (1 - (\gamma_f^* (1 + 2s - 2\gamma_f^*))) \gamma_f^* \right) \right. \\ & \quad \left. + (1+s)(1-d)(1 - (\gamma_f^* (1 + 2s - 2\gamma_f^*))) (1 - \gamma_f^*) \right) \\ & - (\gamma_f^* (1 + 2s - 2\gamma_f^*)) \\ & = 0. \end{aligned} \quad (\text{A.32})$$

Equation A.32 is expanded in a Taylor series in  $s$ . With the assumption that  $s \ll 1$  all terms of the second order or higher are neglected.

$$\frac{1}{2} \frac{\gamma_f^* (d - d\gamma_f^*)}{1 - d + d\gamma_f^*} + 2\gamma_f^* (\gamma_f^* - 1) s + \dots = 0 \quad (\text{A.33})$$

Equation A.33 has the three solutions for the frequency of female gamets at the steady state

$$\begin{aligned} \gamma_{f,1}^* &= 0, \\ \gamma_{f,2}^* &= 1, \\ \gamma_{f,3}^* &= 1 - \frac{4s - d}{4sd}. \end{aligned} \quad (\text{A.34})$$

Since  $0 \leq \gamma_f^* \leq 1$  it is concluded that there exists a polymorphic steady state when

$$\frac{d}{4} \leq s \leq \frac{d}{4-4d}. \quad (\text{A.35})$$

The Equation (A.31) is now used to find the frequency of male gametes carrying allele  $A$  at the steady state

$$\gamma_m^* = \left(1 - \frac{4s-d}{4sd}\right) * \left(1 + 2s - 2\left(1 - \frac{4s-d}{4sd}\right)\right) \quad (\text{A.36})$$

$$= \frac{d-4s+4ds}{8ds} + \frac{4d-16s+16ds}{8d^2} \quad (\text{A.37})$$

With the frequency of gametes with genotype  $A$ , which both males and females contribute with, the frequency of the genotype  $A$  in adult individuals can be calculated.

$$p(A) = \gamma_f \gamma_m + \frac{1}{2} (\gamma_f(1-\gamma_m) + (1-\gamma_f)\gamma_m) \quad (\text{A.38})$$

$$= \frac{3d-12s+12ds}{16ds} + \frac{4d-16s+16ds}{16d^2} \quad (\text{A.39})$$

In this section the conditions for when a polymorphic steady state forms and the frequency the allele  $A$  at this steady state is found. These results agree well with simulations (not shown in this thesis). In the next section the stability of the steady states that have been found is considered.

## A.2.2 Stability analysis of the steady states

The stability of the three steady states found in the previous section is calculated in this section. The first step in the stability analysis is to denote Equation (A.27) and Equation (A.28) from the previous section by

$$\gamma_m(t+1) = f(\gamma_m(t), \gamma_f(t)) \quad (\text{A.40})$$

$$\gamma_f(t+1) = g(\gamma_m(t), \gamma_f(t)) \quad (\text{A.41})$$

The Jacobian matrix for this system at the stationary state  $(\gamma_m^*, \gamma_f^*)$  is

$$A = \begin{pmatrix} \left. \frac{\partial f}{\partial \gamma_m} \right|_{\gamma_m^*, \gamma_f^*} & \left. \frac{\partial f}{\partial \gamma_f} \right|_{\gamma_m^*, \gamma_f^*} \\ \left. \frac{\partial g}{\partial \gamma_m} \right|_{\gamma_m^*, \gamma_f^*} & \left. \frac{\partial g}{\partial \gamma_f} \right|_{\gamma_m^*, \gamma_f^*} \end{pmatrix}. \quad (\text{A.42})$$

The stability of the system can be determined through the eigenvalues of the Jacobian matrix in Equation (A.42). The eigenvalues of the steady state  $(\gamma_{m,1}^* = 0, \gamma_{f,1}^* = 0)$  are:

$$\begin{aligned}\lambda_{1,1} &= 0, \\ \lambda_{2,1} &= \frac{3d - 4}{4(d - s + ds - 1)}.\end{aligned}\tag{A.43}$$

Assuming that both  $s$  and  $d$  are positive and less than 1 it is concluded that steady state ( $\gamma_m = 0, \gamma_f = 0$ ) is stable when  $s > \frac{d}{4-4d}$ . For the next steady state ( $\gamma_{m,2} = 1, \gamma_{f,2} = 1$ ) the eigenvalues is

$$\begin{aligned}\lambda_{1,2} &= 0, \\ \lambda_{2,2} &= \frac{d - 4}{4s - 4}.\end{aligned}\tag{A.44}$$

This steady state is stable when  $s < \frac{d}{4}$ . The final non trivial steady state is analysed in the same way. After the eigenvalues are computed they are expanded in a Taylor series in  $s$ . Since it is assumed that  $s \ll 1$  all non-linear terms are neglected. This gives the following numerically the approxmated eigenvalues

$$\begin{aligned}\lambda_{1,3} &\approx 0.42 + s \left( 2.38 - \frac{4.76}{d} \right), \\ \lambda_{2,3} &\approx 1.22 - s \left( 0.88 - \frac{1.76}{d} \right).\end{aligned}\tag{A.45}$$

There exists a steady polymorphic state when the absolute value of both eigenvalues is less than 1.



# B

## Statistical analysis

### B.1 Statistical analysis of data from laboratory trials

The idea of a model where male fecundity depends on size originates from laboratory experiments performed at the Linnaeus Centre for Marine Evolutionary Biology, University of Gothenburg (Hintz Saltin et al., in press). In this experiment multiple juvenile virgin males are captured and raised. Each matured female is placed in a separate aquarium and assigned two, five or ten males to mate with. Males are kept in aquari for three months and mate frequently with their assigned female. All males are then removed and females are left to produce offspring for a year. The offspring that are produced during this year are genotyped and assigned a father. From this experiment data concerning the size of each male and female and the number of offspring they produced are collected. Using this data Hintz Saltin et al., (in press) found that larger males produced more offspring. This data are examined in this thesis as well in a slightly different way compared to Hintz Saltin et al., (in press). The results found with the statistical analyses preformed in this thesis corresponds well with the result found by Hintz Saltin et al., (in press).

The data that are available from this laboratory trial consist of size of all males and females in the trial and the number of offspring each female and male produced. When using data regarding males only trials that started with ten males are considered in this analysis. The reason for this is to get a higher statistical power. All data that concern females are kept. All males that did not produce any offspring are discarded from the dataset. They are discarded since it can be other factors than size that cause their inability to produce any offspring. Missing observations are also discarded as well as all data that are derived from equations using missing data (for example the calculation of proportional size).

Two different statistical tests are used to conclude if the number of offspring depends on a given trait (see traits below) and the number of produced offspring. For all traits a pooled  $\chi^2$ -test is conducted. An illustration of the  $\chi^2$ -test for the dependency between female size and number of offspring is shown in Figure B.1 and an illustration of the  $\chi^2$ -test for the dependency between male proportional size and the number of offspring is shown in Figure B.2. The null hypothesis that are used in all trials is that the given trait do not affect the number of offspring and thus the number of offspring should only depend on the number of individuals in each group.

The following traits are investigated:

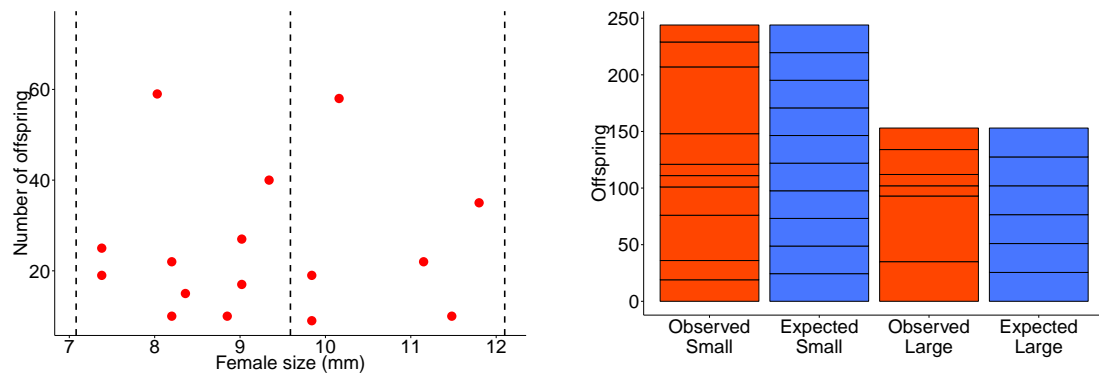
- The size of female size. In this test females are divided into two groups depending on their size, (small and large). This test resulted in a p-value 0.67. This test is illustrated in Figure B.1.
- The number of males the female could mate with. This  $\chi^2$  test gave a p-value of 0.007 where females that had two males produced much fewer offspring than expected and females with five males produced much more than expected. The same test is then done with only the two groups of females that had either five or ten males. This test resulted in a p-value of 0.053.
- The mean size of males the female mated with. Here females are divided into two groups depending on the mean size of the males they mated. This test resulted in a p-value of 0.70.
- The largest male the female mated with. In this test females are divided into two groups depending on the largest male they mated with. This test resulted in a p-value 0.15. Here females that had a largest male, which was smaller, produced more offspring than expected.
- The largest male the female mated with. In this test females are divided into two groups depending on the largest male they mated with. This test resulted in a p-value 0.15. Here females that had a largest male, which was smaller, produced more offspring than expected.
- The size of a male in proportion to other males in the group. In this test males are divided into three groups depending on their size, (small, medium and large). This test resulted in a p-value of 0.031. This test is illustrated in Figure B.2.

This section presented the laboratory trial that were conducted at Linnaeus Centre for Marine Evolutionary Biology, University of Gothenburg by Hintz Saltin et al. (in press). Statistical tests are conducted and as expected it is found that the number of offspring is not independent of size. The next section presents how the data are used to estimate parameters to the model where male fecundity depends on males size that are presented in the main text in Section 2.4.

## B.2 Estimating parameters of the model

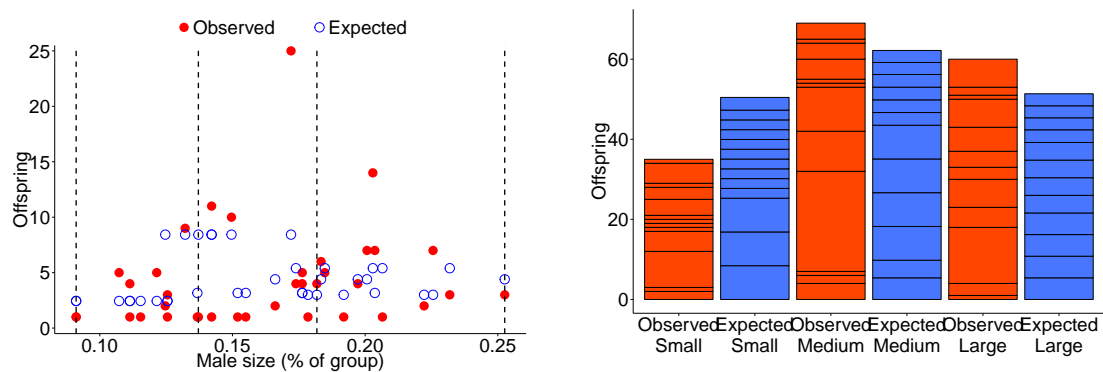
In this section the empirical data from the laboratory trials conducted by (Hintz Saltin et al., in press) are used to estimate parameters to the model where male mating fecundity increases with size (see Section 2.4 in the main text). The model that determines the fecundity of male  $i$  is as follows

$$\phi_i = 1 - d \frac{\hat{\theta}^{(crab)} - \hat{z}_i^{(male)}}{\hat{\theta}^{(crab)} - \hat{\theta}^{(wave)}}. \quad (\text{B.1})$$



(a) Number of offspring over female size. The black dashed lines show the division between small and large females. (b) The total and the expected number of offspring in each group. Each box in the bars represents an individual.

**Figure B.1:** A graphical illustration of the  $\chi^2$ -test for the dependency between female size and number of offspring. The null hypothesis is that the number of offspring is independent of female size. Thus the number of offspring should be equally distributed between large and small females. The  $\chi^2$ -test resulted in a p-value of 0.67.



(a) Number of offspring over male proportional size. The black dashed lines show the division between the three groups in the test. (b) The total number of offspring in each group and the expected number of offspring in each group. Each box in the bars represents an individual.

**Figure B.2:** A graphical illustration of the  $\chi^2$ -test for the dependency between male proportional size and number of offspring. The null hypothesis is that the number of offspring is independent of male proportional size. Thus all males that mated with a specific female should sire the same number of offspring. The  $\chi^2$ -test resulted in a p-value of 0.031.

Here  $\hat{z}_i^{(male)}$  is the phenotype of male number  $i$ ,  $d$  is the fecundity-size parameter  $\hat{\theta}^{(crab)}$  is the optimal phenotype in the habitats that are crab exposed and  $\hat{\theta}^{(wave)}$  is the optimal phenotype in the wave exposed habitat. In this equation there are three unknown parameters:  $d$ ,  $\hat{\theta}^{(crab)}$  and  $\hat{\theta}^{(wave)}$ . The aim is to estimate these parameters with the empirically observed data. The first step is to rewrite Equation (B.1) as follows

$$\begin{aligned}
 \phi_i &= 1 - d \frac{\hat{\theta}^{(crab)} - \hat{z}_i}{\hat{\theta}^{(crab)} - \hat{\theta}^{(wave)}} \\
 &= \frac{d}{\hat{\theta}^{(crab)} - \hat{\theta}^{(wave)}} \left( \frac{\hat{\theta}^{(crab)} - \hat{\theta}^{(wave)}}{d} - \hat{\theta}^{(crab)} + \hat{z}_i \right) \\
 &\propto \frac{\hat{\theta}^{(crab)} - \hat{\theta}^{(wave)}}{d} - \hat{\theta}^{(crab)} + \hat{z}_i \\
 &= \beta + \hat{z}_i.
 \end{aligned} \tag{B.2}$$

Here  $\beta = \frac{\hat{\theta}^{(crab)} - \hat{\theta}^{(wave)}}{d} - \hat{\theta}^{(crab)}$ . The next step is to assume that the mating is a multinomial process where each trial consist of the female in the group selecting one male to mate and produce an offspring with. Here the probability of selecting a given male depends on the size of that male. The probability to observe exactly the same configuration of offspring as in the empirical data is

$$p(\text{observed}|\beta) = \prod_{i=1}^{N(f)} \frac{n_i!}{N_j^{(m)}} \prod_{j=1}^{N_j^{(m)}} p_{i,j}^{x_{i,j}}. \tag{B.3}$$

Here  $N^{(f)}$  is the number of females (same as number of groups),  $N_j^{(m)}$  is the number of males female  $j$  has,  $n_i$  is the number of offspring female  $i$  produced,  $x_{i,j}$  is the number of offspring female  $i$  and male  $j$  produced and  $p_{i,j}$  is the probability for male  $j$  to sire offspring from female number  $i$  under the given model.  $p_{i,j}$  is defined as:

$$p_{i,j} = \frac{\phi_i}{N_j^{(m)} \sum_{k=1}^{N_j^{(m)}} \phi_k} \tag{B.4}$$

Here  $\phi_i$  is defined in Equation (B.2). The value on  $\beta$  that maximize Equation (B.3) is then found. The obtained likelihood as a function of  $\beta$  is shown in Figure B.3. Using the empirical data  $\hat{\theta}^{(crab)}$  is defined as  $mean(\hat{z}_i^{(crab)})$  where  $\hat{z}_i^{(crab)}$  are all males that are larger than the average phenotype, that is all  $\hat{z}_i^{(male)} > mean(\hat{z}_i^{(male)})$ . In the same way  $\hat{\theta}^{(wave)}$  is defined as the mean of all individuals that are smaller than the average phenotype, that is all  $\hat{z}_i^{(male)} < mean(\hat{z}_i^{(male)})$ . With the estimated



values of  $\beta$ ,  $\hat{\theta}^{(crab)}$  and  $\hat{\theta}^{(wave)}$  the value of  $d$  is calculated. The estimated values are  $\beta = 0.186$ ,  $\hat{\theta}^{(crab)} = 5.08$  and  $\hat{\theta}^{(wave)} = 4.08$ . These values give a  $d$  value of 0.194.

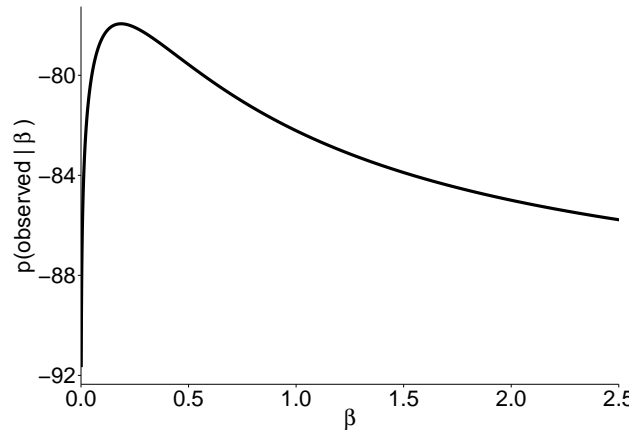
This  $d$  value is evaluated with a pooled  $\chi^2$ -test in a similar way as the previous section. But here the null hypothesis is that the expected number of offspring a given male contributes with is proportional to the fecundity of that male given by Equation (B.1). The null hypothesis could not be rejected with a p-value of 0.085. The  $\chi^2$  is illustrated in Figure B.4.

The system is then transformed so it can be compared to the simulations conducted in this thesis. The system is transformed so that  $\hat{z}_i = z_i * \alpha + c$ . The values of  $\hat{\theta}^{(crab)}$  and  $\hat{\theta}^{(wave)}$  are transformed in a similar way. Giving the follwin equation for  $\phi_i$ ,

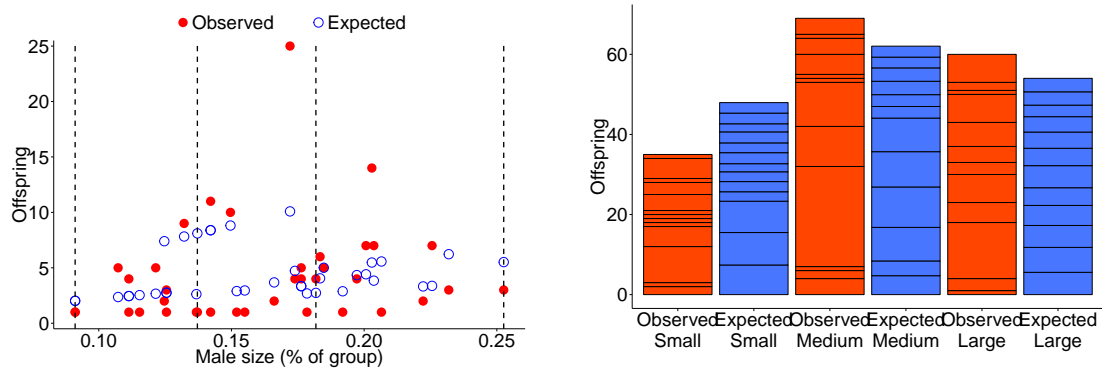
$$\begin{aligned} \phi_i &= 1 - d \frac{\frac{\theta^{(crab)}+c}{\alpha} - \frac{z_i+c}{\alpha}}{\frac{\theta^{(crab)}+c}{\alpha} - \frac{\theta^{(wave)}+c}{\alpha}} \\ &= 1 - d \frac{\theta^{(crab)} - z_i}{\theta^{(crab)} - \theta^{(wave)}}. \end{aligned} \tag{B.5}$$

Equation (B.5) shows that the optimal phenotypes can be transformed in any possible way while the value and the  $d$  value stays the same. The  $d$  value that is found with the empirical data can therefore be used directly in simulations with different values of  $\hat{\theta}$ .

In this section a value for the  $d$  parameter is estimated. It can be concluded that the model presented where male fecundity depends on size can not fully explain the observed data. However this model still gives a better explanation to the data than random mating.



**Figure B.3:** The probability to observe the empirical data under the model in Equation (B.3) as a function of  $\beta$ .



(a) Number of offspring over male proportional size. The black dashed lines show the division between the three groups in the test.

(b) The total number of offspring in each group and the expected number of offspring in each group. Each box in the bars represents an individual.

**Figure B.4:** A graphical illustration of the  $\chi^2$ -test for the dependency between male proportional size and number of offspring. The null hypothesis is that the expected number of offspring is given by the mating model described in equation B.3 with parameters  $d = 0.194$ ,  $\hat{\theta}^{(crab)} = 5.08$  and  $\hat{\theta}^{(wave)} = 4.06$ . The null hypotheses could not be rejected with a p-value of 0.085.