An investigation of sympatric speciation in diploid organisms by computer simulation

by

Robert Puddicombe

Submitted for the degree of Doctor of Philosophy

Department of Computing
Faculty of Engineering and Physical Sciences
University of Surrey
2015

© Robert Puddicombe 2015
Declaration of originality

This thesis and the research to which it refers are the results of my own efforts. Any ideas, data, image or text resulting from the work of others (whether published or unpublished) are fully identified as such within the work and attributed to their originator in the text, bibliography or footnotes. This thesis has not been submitted in whole or in part for any other academic degree or professional qualification. I agree that the University has the right to submit my work to any plagiarism detection service for originality checks. Whether or not drafts have been so-assessed, the University reserves the right to require an electronic version of the final document (as submitted) for assessment as above.

Robert Puddicombe
September 2015
Abstract

This thesis proposes that sympatric speciation can arise, in certain circumstances, in diploid sexually-reproducing organisms by mechanisms other than autopolyplody. This proposition is verified by a series of computer simulations of the natural, biological, genetic processes of reproduction.

The research is essentially a series of logical experiments using a computer simulation of natural genetic mechanisms of reproduction. In this methodology there is no direct comparison with biological data but the simulation follows, as closely as possible, natural genetic mechanisms of reproduction. The selected modelling method uses a framework derived from Penna [1995] which represents individual genes as binary digits on digital chromosomes and incorporates randomised simulations of recombination, copying-errors, meiosis and zygote production.

The literature search identified some difficulties with defining species but a choice was made to use the ‘Biological Species Concept’, which relies on reproductive isolation as evidence of speciation. Sympatric speciation by recombinational processes is said by some to be rare or unknown in nature, but sympatric speciation by autopolyplody is said to be common in plants. The simulations reported here exclude the possibility of autopolyplody and rely only on recombination and mutation processes.

The model described by Penna [1995] was shown to be inadequate for representing multiple species. It was therefore modified to allow more genetic variability but to restrict mating to compatible phenotypes and to allow assortative mating.

Simulations demonstrated that sympatric speciation can occur in scenarios based on the Dobzhansky-Müller model and that these processes can be facilitated by the intervention of selective predators. Another form of sympatric speciation was also detected where paired, dominant and recessive alleles lead to incompatible phenotypes.

No reference has been found in the literature to modelling the Dobzhansky-Müller mechanism, except for Gavrilets [1997] who describes a mathematical analysis rather than a simulation.

**Key words:** diploid, genetics, modelling, speciation, sympatric, assortative mating, predation, Dobzhansky-Müller
Acknowledgements

This research project is financed by the Computing Department of University of Surrey, without whose support it would not have been undertaken. The continued advice, guidance and assistance of Dr André Grüning has also contributed immeasurably to reaching the present stage of the project. Dr Jonathan Clark has also contributed a wealth of biological insight which has helped me to understand the subject area.

I would also like to thank my wife, Elizabeth, for her meticulous help as a proof-reader and her support for this PhD enterprise.
# Contents

1 Introduction .................................................. 16
   1.1 Overview of the research topic ......................... 16
   1.2 The reporting process .................................. 18
   1.3 Layout of this Thesis .................................. 20

2 Literature Review: I - Sympatric Speciation .......... 21
   2.1 Sympatric speciation .................................... 21
   2.2 Sympatric speciation in nature ......................... 26
   2.3 Mechanisms of speciation ................................. 30

3 Literature Review: II - Genetic Modelling .......... 36
   3.1 Modelling objective ..................................... 36
   3.2 Possible modelling techniques ......................... 37
   3.3 The Penna model ......................................... 39
   3.4 Origin and development of the Penna model ........... 43
   3.5 Proposed modelling method ............................... 51

4 Development of the Sympatricia model .............. 52
   4.1 The development process ................................. 52
   4.2 The basic Penna Model ................................ 54
   4.3 Implementation of the Sympatrá Model ............... 56
   4.4 Adding environmental variation ....................... 58
   4.5 Recognising colonies .................................... 60
   4.6 Modelling multiple species .............................. 65
   4.7 Selection mechanisms .................................... 67
5 Simulations using the Sympatria model

5.1 Simulations of the growth in genotype diversity

5.2 Simulations of environmentally acquired mutations

5.3 Simulations maintaining multiple genotypes

5.4 Simulations with the Dobzhansky-Müller mechanism

5.5 Simulations of Dobzhansky-Müller with predators

6 Conclusions

6.1 Review of the methodology

6.2 Conclusions from simulations

6.3 Overall conclusions

6.4 Possible future developments

A Details of the Sympatria model

A.1 Input to the model

A.2 Output mechanisms

A.3 Technicalities of the model

B Theoretical analysis

B.1 Probability of viability in the Penna model

C Simulation data

C.1 Simulations on the growth of genotype diversity

C.2 Simulations on environmentally acquired mutations

C.3 Simulations maintaining multiple genotypes

C.4 Simulations with the Dobzhansky-Müller mechanism

C.5 Simulations of Dobzhansky-Müller with predators
List of Figures

4.1 Environmental death rate \( (D) \) as a function of environmental value \( (V) \) used to change the death rate in different environments. ............................................ 59

4.2 Simplistic examples of \( C_L \) near compatibility boundaries ........................................... 63

4.3 Typical maps of: Genes, Compatibility, and Colonies .................................................. 65

5.1 Growth and decline of diversity for different probabilities of recombination \( (P_R) \). 79

5.2 Purification of mutations from the ends of chromosomes as time progresses ....... 80

5.3 Linear variation of environmental value across the lattice used to test the acquisition of ‘protective’ genes (EVLN). ......................................................... 84

5.4 Mutations acquired in zone \( (0,0) \) of high \( V \) in a benign environment \( (\Delta D_V = 0.005) \). 85

5.5 Mutations acquired in zone \( (0,1) \) of medium \( V \) in a benign environment \( (\Delta D_V = 0.005) \). .......................................................... 86

5.6 Mutations acquired in zone \( (0,2) \) of low \( V \) in a benign environment \( (\Delta D_V = 0.005) \). 86

5.7 Rapid growth of the population to fill the lattice in a benign environment \( (\Delta D_V = 0.005) \). .......................................................... 87

5.8 Typical map of gene types in a stable population \( t = 240 \) in a benign environment \( (\Delta D_V = 0.005) \). ......................................................... 88

5.9 Mutations acquired in a zone \( (0,0) \) of high \( V \) in an aggressive environment \( (\Delta D_V = 0.200) \). .......................................................... 88

5.10 Mutations acquired in zone \( (0,1) \) of medium \( V \) in an aggressive environment \( (\Delta D_V = 0.200) \). .......................................................... 89

5.11 Mutations acquired in zone \( (0,2) \) of low \( V \) in an aggressive environment \( (\Delta D_V = 0.200) \). .......................................................... 89

5.12 Population growth to fill only about 1/3 of the lattice in an aggressive environment \( (\Delta D_V = 0.200) \). .......................................................... 90

5.13 Map of gene types with stable population at \( t = 1800 \) covering about 1/3 of the lattice in an aggressive environment \( (\Delta D_V = 0.200) \). .......................................................... 90
5.14 Compatibility ($C_Q$) over time for uniform environmental value ($V$) .......... 92
5.15 Variation of compatibility ($C_Q$) with gradient ($\Delta V_c$) of the environmental value.
   Case 5.2.3.3 ................................................................. 94
5.16 Gene map at time 15000 with gradient $\Delta V_c = 1.0$ showing sparse population in
   hostile areas. Case 5.2.3.3 ................................................ 94
5.17 Stable gene patterns maintained over a long period using phenotype matching
   (FPT0) ................................................................................. 100
5.18 Longterm genetic diversity from *Uniform Random Complement*, increases with
   phenotype matching tolerance ($T_{pm}$) Recombination $P_R = 20\%$ (FPT0) ..... 101
5.19 Longterm genetic diversity from *Uniform Random Zero*, becomes unstable with
   high phenotype matching tolerance ($T_{pm}$) Recombination $P_R = 20\%$ (FPT0) .. 103
5.20 Genetic diversity over time from *Uniform Random Complement* with continual
   mutation (FPT1) ..................................................................... 105
5.21 Genetic diversity over time from *Uniform Random Zero* with continual mutation
   (FPT1) ................................................................................. 105
5.22 Genetic diversity over time from *Uniform Random Zero* with various AM index.
   $P_R = 20\%$ (FAM0) .............................................................. 108
5.23 Genetic diversity over time from *Uniform Random Zero* with various AM index.
   $P_R = 40\%$ (FAM0) .............................................................. 109
5.24 Genetic diversity over time with PM and AM operating, showing rise in unmutated
   genotype. (PTAM5) ............................................................. 111
5.25 Populations of three main genotypes over time, with PM and AM operating
   (PTAM5) ............................................................................. 112
5.26 Gene-map showing 3 genotypes at 20000 time-steps: RR grey, 00 black, AA red,
   white empty. (PTAM5-007) .................................................. 114
5.27 Gene-map showing only the two original genotypes after 20000 time-steps (DLPA0-005)
   ...................................................................................... 116
5.28 Average genotypes populations against time, with strong PM & AM operating
   ($T_{pm} = 0$ & $q_{AM} = 4$). (DLPA0-000) ................................. 117
5.29 Slow population shift without AM or PM. (DM04-000) .............................. 123
5.30 Rapid population shift with AM only. (DM04-001) ................................. 123
5.31 Slow population shift with PM only. (DM04-002) ................................. 124
5.32 Rapid population shift with AM and PM (DM04-003) ............................ 124
5.33 Population shift to two new species: $P_R = 0.2$ & $q_{am} = 4$ (DM08-000)  
5.34 Population shift without AM applied and no $AB$ exclusion (DM11-000)  
5.35 Population shift with AM applied ($q_{am} = 4$) and no $AB$ exclusion (DM11-001)  
5.36 Population shift with 50% predation rate ($D_P = 0.5$) to exclude $AB$ types but no AM ($q_{am} = 0$). (PR00-000)  
5.37 Population shift with 100% predation rate ($D_P = 1.0$) to exclude $AB$ types but with AM ($q_{am} = 4$). (PR00-001)  
5.38 Two examples of gene maps showing colony development at 500 time-steps with various predation rates ($D_P$) depleting the $aabb$ genotype. Genotypes are colour-coded. (PR01)  
5.39 Mean population of $AAbb$ & $aaBB$ at $t=2000$ for various general death and predation rates (PR18-21)  

B.1 Probability of viable offspring v Genes on Haplotype  
C.1 Zygotic case FZG0 input data  
C.2 Zygotic case FZG1: URC, $P_R = 0.4$ and 50000 time-steps  
C.3 Zygotic case FZG2: URC, $P_R = 0.2$ and 150000 time-steps  
C.4 Zygotic case FZG0 (URC): growth in diversity  
C.5 Zygotic case FZG0 (URZ): growth in diversity  
C.6 Zygotic case FZG0 (RZ): growth in diversity  
C.7 Zygotic case FZG1 & FZG2: growth in diversity  
C.8 Preliminary test of phenotype matching (PT03): input  
C.9 Longterm stable gene patterns (FPT0): results  
C.10 Change in genetic diversity. URC, $P_R = 20\% 40\% 80\%$ (FPT0)  
C.11 Change in genetic diversity. URZ, $P_R = 20\% 40\% 80\%$ (FPT0)  
C.12 Change in genetic diversity. RZ, $P_R = 20\% 40\% 80\%$ (FPT0)  
C.13 Change in genetic diversity. URC, $P_R = 20\% 40\% 80\%$ (FPT1)  
C.14 Change in genetic diversity. URZ, $P_R = 20\% 40\% 80\%$ (FPT1)  
C.15 Change in genetic diversity. RZ, $P_R = 20\% 40\% 80\%$ (FPT1)  
C.16 Change in genetic diversity. URC, $P_R = 20\% 40\% 80\%$ (FAM0)  
C.17 Change in genetic diversity. URZ, $P_R = 20\% 40\% 80\%$ (FAM0)
C.18 Change in genetic diversity. RZ, $P_R = 20\% 40\% 80\%$ (FAM0) ............... 210
C.19 Genetic diversity over time (PTAM5) .................................................. 212
C.20 Average genotypes population $P_R = 60\%$ and $q_{am} = 4$ (DLPA0-000 & -001) ... 215
C.21 Gene-map showing only 2 genotypes at 20000 time-steps (DLPA0-005) ............ 216
C.22 Genetic diversity from initial URC pop. (FUN0) ..................................... 219
C.23 Genetic diversity from initial URZ pop. (FUN0) ..................................... 220
C.24 Genetic diversity from initial RZ pop. (FUN0) ..................................... 221
C.25 DM04-000 Population data ................................................................. 224
C.26 DM04-000(000) Example genotype maps at 5k time-steps ......................... 225
C.27 DM04-001 Population data ................................................................. 226
C.28 DM04-001(000) Example genotype maps at 5k time-steps ......................... 227
C.29 DM04-002 Population data ................................................................. 228
C.30 DM04-002(000) Example genotype maps at 5k time-steps ......................... 229
C.31 DM04-003 Population data ................................................................. 230
C.32 DM04-003(000) Example genotype maps at 5k time-steps ......................... 231
C.33 DM07-000 Population data ................................................................. 234
C.34 DM07-000(000) Example genotype maps at 5k time-steps ......................... 235
C.35 DM07-001 Population data ................................................................. 236
C.36 DM07-001(000) Example genotype maps at 5k time-steps ......................... 237
C.37 DM07-002 Population data ................................................................. 238
C.38 DM07-002(000) Example genotype maps at 5k time-steps ......................... 239
C.39 DM07-003 Population data ................................................................. 240
C.40 DM07-003(000) Example genotype maps at 5k time-steps ......................... 241
C.41 DM07-004 Population data ................................................................. 242
C.42 DM07-004(000) Example genotype maps at 5k time-steps ......................... 243
C.43 DM07-005 Population data ................................................................. 244
C.44 DM07-005(000) Example genotype maps at 5k time-steps ......................... 245
C.45 DM07-006 Population data ................................................................. 246
C.46 DM07-006(000) Example genotype maps at 5k time-steps ......................... 247
C.47 DM07-007 Population data ................................................................. 248
## List of Tables

2.1 Viable genotypes in the Dobzhansky-Müller model .......................... 32

4.1 Hamming Distance compared with genetic compatibility in extreme, simplistic cases ......................................................... 63

5.1 Key Inputs for growth and decline of number of genotypes ........... 79

5.2 Key input parameters for hostile environments ............................ 85

5.3 Key input parameters values for uniform and varying environmental value ... 92

5.4 Compatibility ($C_Q$) over time for uniform environmental value ($V$) (Time is in 1000's). Case 5.2.3.2 .................................................. 93

5.5 Summary of assumptions case PTAM5 ........................................ 110

5.6 Genotype populations at time 20000 (run PTAM5-009) .................... 113

5.7 PM expressed genes (run PTAM5-009) ........................................ 113

5.8 PM mismatched genes (run PTAM5-009) ...................................... 113

5.9 Summary of assumptions case DLPA0 ........................................ 115

5.10 DM04 Use of Phenotype matching and Assortative mating .................. 122

5.11 DM07 Possible gametes ........................................................ 125

5.12 DM07 Influence of AM index ($q_{AM}$) and mating range ($R_m$) on $aabb$ population at time 100000 .................................................. 126

5.13 DM08 Assumptions for testing separated loci for Aa and Bb ............... 128

5.14 DM11 Assumptions for testing AM, without excluding the AB phenotype ... 130

5.15 Population shift with no additional mutations ($P_m = 0$) (DM12-000) .................. 133

5.16 Population shift with 10% probability of mutations ($P_m = 0.1$) (DM12-001) 133

5.17 Population shift with 50% probability of mutations ($P_m = 0.5$) (DM12-005) 133

5.18 Mean population of AAbb & aaBB at $t=2000$ for various general death ($D_0$) and predation rates ($D_P$) (PR18-21) ............................... 143
C.25 PR01 input data ................................................................. 310
C.26 PR18 input data ................................................................. 315
C.27 PR19 input data ................................................................. 316
C.28 PR20 input data ................................................................. 317
C.29 PR21 input data ................................................................. 318
C.30 Mean (AAbb & aaBB) population at t=2000 (PR18-21) ........... 319
Publications


**Puddicombe, R.D.** (2011) *Evolutionary speciation* Proceedings of 8th Annual Computing Department PhD Conference, University of Surrey

**Puddicombe, R.D.** (2011) *Development of colonies of genotypes in a sympatric model of diploid entities* 8th European Conference on Mathematical and Theoretical Biology, Krakow

**Puddicombe, R.D.** (2013) *Modelling of sympatric speciation* PhD Confirmation Report, University of Surrey [unpublished]

Presentations of work

In June 2011, a short presentation on the project up to that point was made at the 8th European Conference on Mathematical and Theoretical Biology held in Kraków, Poland. The presentation included some preliminary results on the increase and subsequent decrease in the number of distinct genotypes of plants in the model under various operating conditions.

Academic Contacts

The principal benefit of the presentation at Kraków was a cordial introduction to Professor Stanisław Cebrat of Wroclaw University, Poland and the resulting opportunity to visit him and his team there in October 2011. This visit was very informative and Prof. Cebrat discussed his recent modelling of age-related disease and full-size human genome in derivatives of the Penna model. Some useful technical ideas on the computational aspects of the model were provided by other members of Cebrat’s research team, especially Wojciech Waga.
### List of Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description of entity</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta(h)$</td>
<td>Function for counting mutations (bits) in haplotype $h$</td>
</tr>
<tr>
<td>$C_Q$</td>
<td>Compatibility of columns at $1/4$ and $3/4$ points of the lattice.</td>
</tr>
<tr>
<td>$C_L$</td>
<td>Local compatibility of a plant.</td>
</tr>
<tr>
<td>$C_N$</td>
<td>Local compatibility of a plant after normalisation.</td>
</tr>
<tr>
<td>$D$</td>
<td>Environmental death rate (i.e. probability of death per time-step)</td>
</tr>
<tr>
<td>$D_P$</td>
<td>General predation death rate</td>
</tr>
<tr>
<td>$D_{Pi}$</td>
<td>Probability of predation of plant $i$</td>
</tr>
<tr>
<td>$D_c(V)$</td>
<td>Environmental death rate as a function of environmental value $V$</td>
</tr>
<tr>
<td>$D_0$</td>
<td>Base environmental death rate</td>
</tr>
<tr>
<td>$\Delta D_V$</td>
<td>Increase in environmental death rate per unit increase in $V$</td>
</tr>
<tr>
<td>$\Delta V_c$</td>
<td>Increase in environmental value per lattice column</td>
</tr>
<tr>
<td>$\Delta V_m$</td>
<td>Increased tolerance to environmental value per beneficial mutation</td>
</tr>
<tr>
<td>$E_i$</td>
<td>Gene pattern expressed in the phenotype</td>
</tr>
<tr>
<td>$E_{Pa}$</td>
<td>Pattern of expressed genes at loci $L_{Pa}$ which attracts predators</td>
</tr>
<tr>
<td>$E_{Pd}$</td>
<td>Pattern of expressed genes at loci $L_{Pd}$ which deters predators</td>
</tr>
<tr>
<td>$F_M$</td>
<td>Proportion of the population of plants with $M$ mutations in the genotype</td>
</tr>
<tr>
<td>$G_{Ai}$</td>
<td>Gamete ‘i’ from parent ‘A’</td>
</tr>
<tr>
<td>$G'_{Ai}$</td>
<td>Complement of gamete ‘i’ from parent ‘A’</td>
</tr>
<tr>
<td>$H_{A0}$</td>
<td>First (haplotype) chromosome of plant ‘A’</td>
</tr>
<tr>
<td>$H_{A1}$</td>
<td>Second (haplotype) chromosome of plant ‘A’</td>
</tr>
<tr>
<td>$[H_p,H_q]$</td>
<td>The two haplotypes in any plant</td>
</tr>
<tr>
<td>$L_{am}$</td>
<td>Loci at which assortative mating is assessed</td>
</tr>
<tr>
<td>$L_c$</td>
<td>‘Critical’ loci at which double mutations are fatal</td>
</tr>
<tr>
<td>$L_L$</td>
<td>Loci which improve tolerance of low environmental value</td>
</tr>
<tr>
<td>$L_H$</td>
<td>Loci which improve tolerance of high environmental value</td>
</tr>
</tbody>
</table>
List of Symbols continued

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description of entity</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_{Pa}$</td>
<td>Loci which attract predators</td>
</tr>
<tr>
<td>$L_{Pd}$</td>
<td>Loci which deter predators</td>
</tr>
<tr>
<td>$L_{pm}$</td>
<td>Loci at which phenotype matching is assessed</td>
</tr>
<tr>
<td>$L_g$</td>
<td>Loci at which genes are reported for population sampling</td>
</tr>
<tr>
<td>$M_{A1}$</td>
<td>Number of mutation bits in haplotype ‘A1’</td>
</tr>
<tr>
<td>$M_T$</td>
<td>Number of double mutations tolerated in the chromosome</td>
</tr>
<tr>
<td>$M_Z$</td>
<td>Number of coincident (fatal) mutations in zygote Z</td>
</tr>
<tr>
<td>$N$</td>
<td>Number of gene loci (bits) in a haplotype</td>
</tr>
<tr>
<td>$P_m$</td>
<td>Probability of one mutation per gamete copy</td>
</tr>
<tr>
<td>$P_R$</td>
<td>Recombination rate (probability of cross-over)</td>
</tr>
<tr>
<td>$P_v$</td>
<td>Probability of a viable offspring</td>
</tr>
<tr>
<td>$(p : q)$</td>
<td>Haplotype of $p$ mutated genes (1) followed by $q$ 'wild' genes (0)</td>
</tr>
<tr>
<td>$q_{am}$</td>
<td>Index which increases the influence of assortative mating</td>
</tr>
<tr>
<td>$R$</td>
<td>A typical random haplotype chromosome</td>
</tr>
<tr>
<td>$t$</td>
<td>Time measured in computing cycles</td>
</tr>
<tr>
<td>$t_{max}$</td>
<td>Length of a modelling run in time-steps</td>
</tr>
<tr>
<td>$T_m$</td>
<td>Maximum number of double mutations tolerated in ageing process</td>
</tr>
<tr>
<td>$T_{pm}$</td>
<td>Number of gene mismatches allowed in phenotype matching</td>
</tr>
<tr>
<td>$T(H_p, H_q)$</td>
<td>Function giving standardized genotype from haplotypes</td>
</tr>
<tr>
<td>$V$</td>
<td>Environmental value</td>
</tr>
<tr>
<td>$V_H$</td>
<td>Upper limit of $V$ above which environmental death rate rises</td>
</tr>
<tr>
<td>$V_L$</td>
<td>Lower limit of $V$ below which environmental death rate rises</td>
</tr>
<tr>
<td>$w_c$</td>
<td>Width of a central block of columns in lattice</td>
</tr>
<tr>
<td>$Z$ or $0$</td>
<td>The zero (unmutated) chromosome</td>
</tr>
<tr>
<td>$\bar{Z}$ or $\bar{0}$</td>
<td>The fully mutated chromosome; the complement of zero</td>
</tr>
<tr>
<td>$\cap$</td>
<td>AND operator on binary numbers</td>
</tr>
<tr>
<td>$\cup$</td>
<td>OR operator on binary numbers</td>
</tr>
<tr>
<td>$\oplus$</td>
<td>Exclusive OR operator on binary numbers</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

This introduction gives an overview of the research topic and outlines the structure of this thesis.

1.1 Overview of the research topic

The focus of this research is establishing what mechanisms of genetics and selection are required to achieve sympatric speciation. This is attempted using a computer model of the processes of genetic reproduction.

The model is not intended to represent any particular species or group of species but is intended to simulate the general evolutionary mechanisms of diploid sexually reproducing organisms.

1.1.1 Speciation

The evolution of species in the natural world is now well established as a scientific reality, succinctly summarised by Coyle [2009]. It is also true that the term species has been used for several centuries\(^1\) and one might assume that it is a well understood concept but, as shown below in the Literature Review, there are a number of difficulties with defining a species. There are consequently problems identifying speciation which is the evolutionary process of separation into species.

Speciation in the natural world is generally divided according to the circumstances in which it occurs, namely: *allopatric, parapatric, peripatric and sympatric*, which signify habitats which are respectively: divided, adjoining, separated or unseparated. The first, and to some extent the second and third, of these situations seem very likely to result in speciation because a group

\(^1\) Oxford English Dictionary OED [2013] cites the earliest biological usage as: 1608 E.Topsell Hist. Serpents ‘Some have taken the word Crocodilus for the Genus, and the severall Species, they distinguish into the Crocodile of the Earth, and the water.’
of individuals in a new habitat with different survival pressures is likely to undergo specialised changes and possibly emerge as a new species incompatible with the original population. It is more interesting to contemplate the mechanisms and circumstances which would result in the emergence of a new species within an existing population during sympatric speciation. Sympatric speciation therefore seems worthy of further investigation and is the subject of this research project.

Sympatric speciation is said, by some authorities (e.g. Coyle and Orr [2004]) to be rarely if ever found in animals, although it can arise in plants. Sympatric speciation in plants often occurs by a process called autopolyplody in which whole chromosomes are duplicated in error resulting in a completely new form which is frequently reproductively incompatible with its parent population. Isolation after autopolyplody is not surprising and therefore the current research does not consider it but looks only at speciation by other genetic processes.

1.1.2 Modelling

In order to investigate the genetic and selection processes around sympatric speciation it is necessary to set up a modelling environment which simulates, as closely as possible, the natural processes of genetics, inheritance and survival. Using a computerised model rather than biological field data allows a wide range of artificial situations to be tested through many thousands of generations to see if speciation occurs. The methodology used in this project therefore aims to simulate the reproductive processes of natural biology as closely as possible in order to test possible mechanisms of speciation.

In order for the model to provide a valid simulation of biological reproduction it needs to do more than simply use evolutionary algorithms. Evolutionary algorithms can be used in a variety of ways but their processes for genetic inheritance are not necessarily the same as in natural reproduction. The literature shows that crossing of genes for offspring may be executed in a variety of ways to achieve evolutionary improvements but these do not always correspond to natural inheritance. For the current project to be successful the modelling algorithms must be selected carefully. The selected model will need to employ algorithms which simulate, as closely as possible, the genetic processes of reproduction in the natural world, or at least in a defined sub-set of the natural world.

The selection of the modelling method and its validity are discussed in Chapter 3, which is Part II of the Literature Review. Chapter 4 goes on to describe the details and development of the selected model.

By using this modelling approach the research project has become a series of logical experiments using the model to apply certain logical rules of genetic inheritance in different circumstances to see if speciation arises. It is not intended to represent any particular organisms or groups of species, but it looks at the possible genetic outcomes of diploid sexual inheritance.
1.1.3 Recognising speciation

Simulations with the potential to achieve speciation are only productive if a working criterion is established for recognising when speciation has occurred. The Literature Review identifies that within the limited context chosen for this project, reproductive isolation is the key indicator of speciation. That is to say: individuals which are reproductively incompatible are assumed to belong to separate species. In this context reproductive incompatibility is evaluated as the ability to produce the next generation without regard to future fertility. This project offers two approaches to identifying reproductive incompatibility and consequently speciation.

Firstly incompatibility may be identified by algorithms which tested the potential reproductive compatibility between individuals and marked out boundaries between incompatible groups. Devising these algorithms was an innovative part of the project as they are more sophisticated than methods employed by others using the same type of model.

Secondly, mechanisms may be devised within the model which exclude compatibility between certain genotypes. These genotypes are then, by definition, reproductively isolated. The development and propagation of these genotypes can then be studied. This method is based on the Dobzhansky-Müller model.

The only reference found in the literature to modelling of the Dobzhansky-Müller mechanism is in Gavrilets [1997], but this was a mathematical analysis rather than a simulation. It therefore appears that the current project may be unique in modelling sympatric speciation by the Dobzhansky-Müller mechanism.

1.1.4 Objective of the research project

The objective of this research project can therefore be summarised as: using computer simulation of biological genetic processes to find the conditions in which sympatric speciation can occur without polyploidy.

This objective will depend on three key tasks:

- defining sympatric speciation
- modelling biological genetic processes
- recognising when speciation occurs

1.2 The reporting process

This research is reported in stages which broadly follow the chronological development of the research programme. The first is the Literature Review in two parts: Chapter 2 investigates
biological aspects of speciation and Chapter 3 looks at modelling techniques. Following the
literature review, the selected modelling method is discussed in detail and the development of a
suitable computer model is described in Chapter 4. A series of simulations which use the model
to investigate speciation processes is described in Chapter 5. From these simulations some
conclusions are drawn in Chapter 6 and some possible future lines of research are identified.

1.2.1 Appendices

Appendix A relates to the model described in Chapter 4, and gives more technical detail about
how it was implemented and controlled. Appendix B shows the mathematical derivations of
some relationships which are discussed in the thesis. The largest Appendix is C which records
details of the input and output data relating to the simulations discussed in Chapter 5. Where
necessary data is reproduced in Chapter 5 in order to clarify the discussion of each simulation.
However, full input data for all the simulations is recorded in the sub-sections of Appendix C.

1.2.2 Limitations on data

This type of genetic modelling produces large volumes of data and all the detail cannot be
reproduced in this document. Therefore, in general, averages, trends and graphs are used to
illustrate the points brought out in each simulation.

The detailed coding of the model, which was written in C++ for this project, is also not
reproduced in this document but the model and all the data recorded during the project are
available from the author.
1.3 Layout of this Thesis

This thesis is laid out in chapters as follows:

- 1: Introduction
- 2: Literature Review: I - Sympatric Speciation
- 3: Literature Review: II - Genetic Modelling
- 4: Development of the *Sympatria* model
- 5: Simulations using the *Sympatria* model
- 6: Conclusions

There are also appendices:

- A: Details of the *Sympatria* model
- B: Theoretical analysis
- C: Simulation data
- Bibliography
Chapter 2

Literature Review: I - Sympatric Speciation

This literature review has been divided into two parts:

- I Sympatric speciation
- II Genetic modelling

The first part reviews the biological background and definitions of sympatric speciation; the second looks at past methods which could be used to model speciation and reviews the *Penna Model* in detail.

The essential combination of these two strands of research into a useful project requires the development of a model which is likely to simulate sympatric speciation and, more importantly, the development of a testing technique to ascertain whether speciation has occurred. These are discussed in Chapter 4, on the development of the modelling method.

2.1 Sympatric speciation

This section researches definitions of *sympatric speciation* which implies the separation of populations of organisms into species in an undivided environment. “Sympatric” has a straightforward definition which presents no ambiguity, but “speciation” is a more problematical concept. It is therefore necessary to find a definition for it which is adequate within the context of the current project but this may not be universally applicable in the natural world.

The word “sympatric” derives from the Greek *sym-* (∑µι) meaning “the same” and *patra* (πατρα) literally meaning “fatherland” but taken to imply country or region OED [2013]. In the context of speciation this is distinguished from *allopatric, peripatric and parapatric* which occur in regions which are physically divided, separate or adjacent, respectively.
Speciation, meaning separation into species, is an altogether more difficult term. The key issues in defining speciation are to define what separates species and how that separation can be identified. In the context of the current research, which operates at a genetic level, it is necessary to find a test for speciation in the genotype rather than in the developed phenotype of the organism. That is the subject of the following section.

2.1.1 Definitions of Species

Before embarking on the modelling of speciation it is important to establish a definition of what it is. The following paragraphs demonstrate that this is not a simple task and a compromise is required so that a working definition may be chosen which is sufficient for the purposes of the current research project.

Historically biologists have used various criteria to establish that species are separate. De Queiroz [2007] lists twelve characteristics of species used by various authorities to define separation. However, as demonstrated below (see 2.1.1.2), these criteria cannot all be applied to all types of organisms and biologists have, in the past, used intuitive ideas to distinguish species.

In chapter 2 of *The Origin of Species*, Darwin [1859], commenting on the definition of species, writes “No one definition has satisfied all naturalists; yet every naturalist knows vaguely what he means when he speaks of a species.” Coyle and Orr [2004] suggest that Darwin “... made little distinction between speciation and adaptation”. It could be considered ironic that in his book entitled *On the Origins of Species* Darwin fails to satisfactorily identify the origin of species. This irony was apparently not lost on Dobzhansky who called his 1951 book *Genetics and the Origins of Species*, a clear reference to Darwin’s title with perhaps the implication that Dobzhansky had succeeded in defining the origin of species where Darwin had not. It may be the crux of the current research to investigate when adaptation becomes speciation.

2.1.1.1 The Biological Species Concept

A well established definition of species, known as the Biological Species Concept (BSC), uses sterile interbreeding as the defining factor. That is to say, organisms which cannot interbreed successfully are not of the same species. Coyle and Orr [2004], p28 attribute the BSC to Dobzhansky and state that “A species is a group of individuals fully fertile inter se, but barred from interbreeding with other similar groups by its physiological properties (producing either incompatibility of parents, or sterility of hybrids, or both).” Mayr [1996] writes of the BSC, “The concept I have just developed is articulated in the so-called biological species definition: ‘Species are groups of interbreeding natural populations that are reproductively isolated from other such groups.’ The isolating mechanism by which reproductive isolation is effected are [sic] properties of individuals. Geographic isolation therefore does not qualify as an isolating mechanism.” Stearns and Hoekstra [2000] attribute the formal definition of the BSC to Dobzhansky (1934)
and record that it was more widely disseminated by Mayr in the 1960’s. Dobzhansky may
have derived this definition from principles current in the 19th Century. Darwin [1859] writes
in chapter 8, “The general view entertained by naturalists is that species, when intercrossed,
have been specially endowed with the quality of sterility, in order to prevent confusion of all
organic forms.” This is an observation of an idea current at that time which reverses the
modern causality by supposing that sterility is ‘designed’ to preserve the isolation of species.
However, Darwin continues “I hope, however to be able to show that sterility is not a specially
acquired or endowed quality, but is incidental on other acquired differences.” This is not an
explicit expression of the BSC but appears to indicate that Darwin tacitly accepted sterile
interbreeding as the characteristic which identifies separated species rather than a mechanism
for protecting the isolation of species. The assumption that sterile interbreeding and species
separation are interlinked can be found in several sources. Darwin also writes, of hybrid crosses
between species; “… the sterility of hybrids could not possibly be of any advantage to them,
and therefore could not have been acquired by the continued preservation of successive profitable
degrees of sterility.” Charlesworth and Charlesworth [2003] similarly discuss the concept when
they write, nearly 150 years later, “There is no selection to maintain compatibility of mating
behaviour between individuals from geographically or ecologically separated populations, or to
maintain harmonious interactions that allow normal development, between genes that have come
to differ in different populations. Like other characteristics that are not subject to selection to
maintain them (such as the eyes of cave-dwelling animals), the ability to interbreed degenerates
over time.” These views may seem contradictory but are compatible and complementary:
Darwin states that there is no evolutionary advantage in building sterility barriers between
groups, whereas Charlesworth and Charlesworth [2003] note that there is no mechanism for
maintaining compatibility of development between separated groups.

It might therefore be assumed that sterile interbreeding is a sufficient definition of speciation,
but this definition is immediately seen to fail because it does not deal with the classification
of organisms which do not sexually reproduce. This group, far from being obscure, probably
constitutes the majority of living organisms, both in numbers of species and of individuals,
including some bacteria, many plants and some animals. These organisms have to be classified
into species on the basis of their observed structure. Charlesworth and Charlesworth [2003] note
that asexually reproductive species are defined by phenotype characteristics (i.e. characteristics
of the developed organism) or more recently by DNA sequencing. Leaving aside the problem
with asexual organisms there are other apparent difficulties with defining speciation by sterile
interbreeding.

2.1.1.2 Revised Definitions of Speciation

Stearns and Hoekstra [2000] go so far as to write “Speciation may be a false concept”. They
seem to suggest that because there are grey areas at the boundaries of some species the BSC
definition of speciation as ‘sterile interbreeding’ is somehow inadequate. De Queiroz [2007] presents an extensive review of the subject and lists twelve factors considered by various authorities as indicative of speciation. He notes that in the course of divergence species develop distinct characteristics in several areas which can be paraphrased as: structural differences, reproductive sterility, incompatible developmental systems, differing mate recognition systems and distinctive ecologies. He continues, “The problem is that these changes do not all occur at the same time, and they do not even necessarily occur in a regular order. ... These properties, attributes such as phenetic (structural) distinguishability, reciprocal monophyly (common ancestry), pre- and post-zygotic reproductive isolation, and so forth, are all properties that lineages acquire as they separate and diverge from one another and therefore provide evidence of lineage separation and divergence.” He proposes that while these changes in key areas are indicators of speciation they do not, individually or jointly, provide a sufficient definition of speciation. He notes that “…some people will draw the cut-off relatively early in the process of divergence, perhaps where differences in quantitative characters make the lineages phenetically (i.e. structurally) distinguishable. Others will draw the cut-off somewhat later, perhaps where the lineages develop an intrinsic reproductive barrier.” De Queiroz considers the various characteristics of divergence as “contingent properties: properties that species may or may not acquire during the course of their existence. In other words, lineages do not have to be phenetically (i.e. structurally) distinguishable, diagnosable, monophyletic (i.e. with a common ancestor), intrinsically reproductively isolated, ecologically divergent, or anything else to be considered species.” While insisting that neither sterile interbreeding nor any other characteristic is a sufficient indicator of speciation, he proposes defining species as “segments of separately evolving metapopulation lineages” but does not adequately define ‘separately evolving metapopulation lineages’, simply saying that they “only have to be evolving separately from other lineages.” This seems to require some further explanation of what is meant by ‘evolving separately’ and ‘lineage’.

### 2.1.2 Species Boundaries

Some species appear to be, at the present time, in a semi-detached relationship with other species with which they partially interbreed. This might seem to confuse the definition of the species boundary. Taylor et al. [1997] quote the example of crows in northern Britain. The Hooded Crow (Corvus cornix) inhabits a north-western territory including the Scottish Highlands whereas the territory of the Carrion Crow (Corvus corone) covers parts of northern England but overlaps with the area of its hooded relative. It appears that the two groups interbreed in the overlap but the interbred offspring are less successful, in evolutionary terms less ‘fit’, than pure-bred birds from either group. Consequently a degree of separation of the species is maintained. This is an example of species interbreeding but remaining to some extent distinct. In a similar case quoted by Charlesworth and Charlesworth [2003], a species of fruitfly (Drosophila pseudoobscura) in the south-western USA and Central America is found to breed
semi-fertilely with a population in Bogatá, Colombia. The female hybrid offspring are fertile but the males infertile. Importantly, Charlesworth & Charlesworth continue “There is therefore no compelling reason to consider it (i.e. the Bogatá population) as a separate species, although it is starting to develop reproductive isolation, as indicated by the sterility of the hybrid males.” This statement indicates that Charlesworth & Charlesworth accept the proposition that separation of species is synonymous with reproductive isolation. Charlesworth & Charlesworth also cite the ‘Monkeyflower’ of the north-western USA, *Mimulus lewisii* which has pink flowers and *Mimulus cardinalis* with red flowers. These are pollinated, respectively, by bees and hummingbirds. The pink flowers, which appear brighter to bees, have developed a wider landing platform for their insect pollinators, while the red flowers, which appear darker and less attractive to bees, are larger, narrower and tubular with abundant nectar of low sugar-content, all of which suit the hummingbirds. Despite these structural divergences to satisfy their respective pollinators, the plants can be artificially crossbred successfully, producing healthy and fertile hybrids. These plants can be said to be ‘reproductively isolated’ but are not mutually infertile, which makes it unclear whether these plants are separate species. It is incidentally noteworthy that plants do not actively select their sexual partners as animals might: here the pollinators make the selection. Nevertheless, the evolution of specialised attractive structures is acting as a proxy for sexual selection in tending to separate two otherwise compatible groups.

Geographical factors indicate that the incomplete isolation of the Bogatá fruitflies has developed in about 200,000 years, which would have involved several million generations. A much more rapid speciation process is also identified by Charlesworth and Charlesworth [2003] in the cichlid family of fish (*Cichlidae*) in Lake Victoria, East Africa. By referring to them as ‘species’, Charlesworth & Charlesworth imply, but do not state, that there is no residual reproductive compatibility between the fish species; this implies that there was no longer any interbreeding, because the case would have no special significance and would not be worth reporting if there was residual interbreeding. Geology indicates that the lake has existed for only 14,600 years during which time it appears that sexual selection for colouration has driven the fish into 500 identified ‘species’ indicating, according to Charlesworth & Charlesworth, that speciation has occurred on average every 2000 to 3000 years. As well as being useful examples of partial or complete speciation, these cases point to an implicit assumption by these authorities cited that speciation will be achieved only when a state of sterile interbreeding is reached. Charlesworth & Charlesworth state “… typically, several tens of thousands of years seem to be needed for a new species to be formed,” and “Given enough evolutionary divergence, complete reproductive isolation seems inevitable.” These authors clearly assume ‘reproductive isolation’ to be the key criterion for judging speciation. Nevertheless the ‘reproductive isolation’ definition is not universally accepted.
2.1.2.1 Conclusion on the definition of speciation

The examples of partially separate species quoted above have in common an assumption, stated or implied, that sterile interbreeding is the de facto benchmark by which speciation is identified. The uncertainty at the boundaries for some species which currently partially interbreed does not undermine this assumption but points to the use of sterility of interbreeding to ascertain when speciation has finally occurred. ‘Sterile interbreeding’ therefore appears to be the best criterion available for judging speciation in sexually reproductive organisms. Using this definition leaves the monkey-flowers, crows, and fruitflies quoted in 2.1.2, as unseparated species. They could be considered as lying on a continuum, in that order, between total species unity and total species separation: the flowers do not interbreed but could do; the birds do interbreed but with some disadvantages; and the flies interbreed with major disadvantages. Totally sterile interbreeding remains a valid marker for total species separation. This continuum of partial interbreeding provides a useful framework for impending species separation when modelling evolution. Groups of cross-bred individuals becoming less successful in some aspect of fitness may be a useful indicator of impending speciation. In the context of the current research project, the model has been set up to deal only with sexual reproduction and therefore reproductive isolation is a workable definition of the separation of groups of organisms. Whether or not this criterion exactly aligns with the biological definition of speciation is a semantic question, but for the purposes of the current research reproductive isolation can be used as a satisfactory measure of the separation of groups within a population.

The need to decide upon a workable definition of speciation is succinctly summarised by Brookfield [2002] quoted by Coyle and Orr [2004] as writing: “… the ‘species problem’ is not a scientific problem at all, merely one about choosing and applying a convention about how we use a word. So, we should settle on our favourite definition, use it, and get on with the science.”.

For the purposes of the current project reproductive isolation is taken as the indicator of speciation and in this context reproductive isolation is assumed to be indicated by the immediate infertility between parents without making any assessment of the fertility of subsequent generations. This assumption is an approximation which excludes the possibility of infertility developing progressively over several generations.

2.2 Sympatric speciation in nature

Having reviewed how a species separation might be defined, a remaining question is whether sympatric speciation is an observed process in nature. There are various views on the possibility of sympatric speciation.
2.2.1 Rarity of sympatric speciation

In their book entitled *Speciation*, Coyle and Orr [2004] conclude that there is scant biological evidence for sympatric speciation in the natural world. They present extensive detail (p147) about the speciation of Cichlidae fish in East Africa, as mentioned in Section 2.1.2, above. They show that there is considerable doubt about the unexpectedly fast evolution of these species and point to the possibility of allopatric speciation in small residual pools which may have existed when Lake Victoria was otherwise believed to have been dry 14,600 years ago. They also raise the unanswered question of whether the many groups of differently coloured fish are closely enough related to indicate recent speciation, suggesting that they may be derived from older separate lineages. They conclude (p178), “While we can point to a few promising cases, they do not add up to strong support for the idea that this process is common. While additional work may provide compelling evidence, it is hard to see how the data at hand can justify the current wave of enthusiasm for sympatric speciation”.

By contrast, Jiggins [2006] in his 2006 review of the subject suggests that although sympatric speciation may be difficult to identify it is not necessarily uncommon.

2.2.2 Speciation by polyploidy

Having dismissed sympatric speciation as unlikely, Coyle and Orr [2004] subsequently offer a separation of speciation mechanisms saying (p321) “.. there are two main types, polyploid speciation and recombinational speciation. Polyploid speciation is common ... . Recombinational speciation is of unknown frequency, [and] is less well understood...”. They then concede “... polyploid speciation is instantaneous, [and] sympatric ...” and “... polyploidy is common in plants but rare in animals,” which may be the reason for their separation of polyploidy from their earlier discussion of recombinational speciation in animals.

Polyploidy, the rare duplication of all the chromosomes caused by an error in meiosis, can cause ‘instantaneous’ reproductive isolation from the parent group because the offspring produces incompatible forms of gamete or infertile hybrids. For example, if two diploid parents produce diploid gametes these will create a tetraploid offspring which will then generally produce diploid gametes. These diploid gametes are likely to be incompatible with the ancestral haploid gametes or, if they do cross, will produce a triploid hybrid which (as in the case of the banana *Musa acuminata* and *Musa balbisiana*) cannot produce any gametes because the division of chromosomes during meiosis anaphase I becomes asymmetrical and fails.

Triploid organisms are not always infertile as shown by the ‘edible frog’ (*Pelophylax esculentus*) which is a triploid hybrid between *Pelophylax lessonae* and *Pelophylax ridibundus* which are genetically ‘conventional’ being somatically diploid producing haploid gametes. The somatically triploid *P. esculentus* produces diploid and haploid gametes carrying chromosomes from either
or both parent species, *P. lessonae* and *P. ridibundus*. These can combine with haploid gametes from the parent population in various combinations some of which survive to reproductive maturity and perpetuate the hybrid population. The combinations of gametes and their sexual significance are beyond the reach of the current thesis but are described in detail by Christiansen [2009]. The recent (5Mya) separation of the species *P. lessonae* and *P. ridibundus* is described by Ragghianti et al. [2007] who refer to the species as *Rana ecaulentus* whereas Frost et al. [2006] shows *Pelophylax* as a separate section of the genus *Rana* and Christiansen [2009] uses the *Pelophylax* nomenclature.

Christiansen’s work was on live specimens and he presents detailed analyses of the populations over 70 generations. It would be interesting to simulate these interactions in a computer model, but this is outside the scope of the current research project. The model used in the current research project does not allow for the production of polyploid hybrids and therefore only speciation by recombinational processes is being considered at present.

Coyne & Orr’s 2004 observation that recombinational sympatric speciation is rare does not exclude the possibility that some occurrences of sympatric speciation might exist somewhere in the natural world. This implies that it should be possible to simulate a situation in which sympatric speciation occurs. The current research is based on this assumption.

### 2.2.3 Other speciation observations

Ortiz-Barrientos and Rieseberg [2006] describe two instances where they report sympatric speciation: *Cichlidae* in a crater lake in Nicaragua, and types of palm plant on Lord Howe Island, a remote island between Australia and New Zealand. The fish, *Amphilophus citrenelus* and *Amphilophus zaliosus*, forage in different water columns which is another example of questionable *sympatria* where distinct environments exist side-by-side. The palms *Howea forsteriana* and *Howea belmoreana* flower at different times of year which may be caused by growing in different soil types, volcanic or calcareous.

Jiggins [2006] refers to the same natural examples concerning whether or not sympatric speciation is possible and acknowledges that proving it is difficult, saying “The main reason why the debate over the role of geography in speciation has not been resolved is that distinguishing the alternatives in any particular case is extraordinarily difficult.” He goes on to give “. . . two recent studies that support sympatric speciation.” These again concern the *Cichlid* fish in Nicaragua, and the palm trees on Lord Howe Island, and Jiggins concludes that “In both situations, therefore, there is evidence for the kind of divergent ecological selection that we expect to be associated with sympatric speciation.”

However, the mechanisms of these speciations appear to be: a difference in feeding habits indicated by jaw structure of the fish, and a difference in response to soil acidity in the case of
the palms. The latter case raises a question as to what is truly sympatric: does a difference in soil acidity amount to allopatry in separate areas with different soil types?

2.2.4 Stages in species separation

Coyle and Orr [2004] (p130) confirm that "... models that start with ecological divergence can allow long-term coexistence of sympatrically formed taxa, permitting ample time for the evolution of other isolating barriers." This supports the idea that ecological effects, which in this context could be taken to mean the interaction of the organisms and the environment, can be used to generate the initial separation. Having achieved this, other factors may evolve to reinforce the separation. It therefore appears possible to achieve sympatric speciation but a philosophical question remains as to whether a region with distinctly different environmental characteristics in certain areas is truly sympatric.

Gavrilets [2003] presents the more encouraging suggestion that adaptation alone will not achieve sympatric speciation but that some assortative mating, for example sexual selection, is required to reinforce the separation of nascent species. Gavrilets concludes, "Sympatric speciation is promoted if costs of being choosy\(^1\) are small (or absent) and if linkage between the loci experiencing disruptive selection and those controlling assortative mating is strong." It therefore seems that adaptation of genes, alone, is insufficient to produce speciation and some appropriate selection mechanisms are therefore required in the current research model.

2.2.5 Assortative mating

It appears that sympatric speciation needs a sexual selection pressure initially to maintain the exclusivity of a new hybrid lineage. The selection of a particular type of mate from within a group of available mates is often referred to as assortative mating. Referring to the potential development of a species by hybridization of two butterfly species, Heliconius melpomene and H. cydno, Melo et al. [2009] say: "These hybrids show assortative mating preferences, showing a strong preference for their own color pattern\(^2\) over that of either parental species. This is consistent with a genetic basis to wing pattern preference and implies, first, that assortative mating preferences would facilitate the initial establishment of a homozygous hybrid color pattern by increasing the likelihood that early generation hybrids mate among themselves. Second, once established such a lineage would inherit assortative mating preferences that would lead to partial reproductive isolation from parental lineages."

\(^1\)In this context 'being choosy' appears to mean 'being selective of mates'.

\(^2\)It is interesting to speculate how butterflies can select mates with similar wing patterns to their own as they may not be able to see their own wings and, having emerged from an unattended egg, have no imprinted image of the pattern of their parents' wings.
2.2.6 Research objective concerning sympatric speciation

Melo et al. [2009] offer a summary which might be used to define the current research project when they write, “Homoploid hybrid speciation is the establishment of a reproductively isolated lineage through hybridization without a change in chromosome number. It is therefore distinct from the numerous examples of hybrid speciation through polyploidization in plants.”

This indicates that the current research project should take on the objective of trying to establish whether sympatric speciation is possible without resorting either to environmentally induced allopatry or polyploidy.

2.3 Mechanisms of speciation

2.3.1 Darwin’s difficulty

In his 1996 summary of the development of early 20th century ideas for a genetic definition of speciation, Orr [1996] recounts what he calls “Darwin’s Paradox.” This is the difficulty of finding an irreversible mechanism for a genotype to evolve from a form which is compatible with its ancestors to a form which is reproductively isolated from them, so that a new species is formed. According to Orr, achieving this change at only one gene locus is difficult.

If one considers a diploid organism which exists with two viable variations: one having alleles AA at a particular locus, and an alternative with aa at the same locus, but the intermediate form Aa is non-viable. The problem, as Orr states it, is that to evolve from AA to aa requires at least one intermediate generation with the non-viable form Aa. In terms of a fitness landscape, AA and aa are ‘peaks’ of fitness and Aa is a ‘valley’ between them. The difficulty arises because if Aa is to prevent evolution from species aa back to AA then it will also prevent the initial evolution from AA to aa. It is worth noting that Orr does not address the possibility of the simultaneous mutations in both gametes so that AA changes directly to aa. One must assume that the probability of this occurring is so low as to be considered impossible.

Describing this as Darwin’s Paradox is to some extent an exaggeration because, apart from it not being a paradox, there is no evidence that Darwin was aware of the modern principles of genetics. Mendel published his work on inheritance in 1865, six years after Darwin’s Origin of Species, but Mendel’s work was not widely recognised until the end of the 19th Century after Darwin’s death. Therefore Mendel’s ideas were not combined with the concept of Darwinian evolution until the early 20th Century. The combination of evolution and genetics, often referred to as the Modern Synthesis, is attributed to Dobzhansky in his 1937 book Genetics and the origin of species.

---

3This is not strictly a ‘paradox’ which is defined as “a statement or tenet contrary to received opinion or belief” OED [2013]. It is simply a ‘problem’, or to provide alliteration a ‘difficulty’.
Darwin would not therefore have seen this problem in the same terms as Orr does but there is evidence that Darwin saw the need for a better theory of inheritance to account for evolutionary change. Darwin would have understood that the simple mixing of inherited characteristics in an analogue model, as if they were coloured paints, would tend to produce a uniform, mean, ‘grey’ set of characteristics after many generations, but the radical concept of exchanging digital genes, as described by Mendel, was apparently unknown to him\(^\text{4}\).

Nevertheless, whether Darwin understood the problem or not, it leaves a valid question as to how irreversible genetic changes might occur, and the Dobzhansky-Müller model offers a possible solution.

### 2.3.2 The Dobzhansky-Müller model

“Darwin’s Paradox” which Orr describes in [Orr, 1995, 1996] can be overcome with a mechanism known as the *Dobzhansky-Müller model*, attributed by Coyle and Orr [2004] to Dobzhansky in 1934. This proposes a two-stage genetic change involving at least two loci.

Assume possible alleles \(A\) and \(a\) at one locus and \(B\) and \(b\) at another, where \(A\) and \(B\) are mutations of \(a\) and \(b\), and where the coincidence of alleles \(A\) and \(B\) is non-viable or significantly debilitating.

Consider a population containing only genotype \(aabb\). After acquiring one mutation at each locus the population could additionally include genotypes \(Aabb\) and \(aaBb\). From this mixed population, simply by recombination, the additional genotypes \(AAbb\) and \(aaBB\) could arise. These homozygous forms are mutually incompatible as any cross between them will always contain allele \(A\) with \(B\) which is non-viable. Consequently it can be said that \(AAbb\) and \(aaBB\) are separate species. The separation would become more complete if the ancestral genotypes \(aabb\), \(Aabb\) and \(aaBb\) were to become extinct. If this occurred then \(AAbb\) and \(aaBB\) would then become completely genetically isolated.

Thus the Dobzhansky-Müller model demonstrates that, in theory, it is possible to achieve speciation without the need for a physical barrier or autopolyplody. Sympatric speciation therefore appears to be possible.

Table 2.1, below shows the progressive development of viable genotypes in the model over several generations.

The changes indicated above as \(A\) and \(B\) could involve groups of several genes and could occur sequentially rather than simultaneously, but nevertheless the overall process would remain valid.

\(^{\text{4}}\)There is documentary evidence Galton [2009] that Mendel owned and annotated a German translation of Darwin’s work [1859] but only circumstantial evidence that Darwin may have seen but not appreciated the significance of Mendel’s 1865 work *Versuche üuber Pflanzen-Hybriden*. Darwin was able to understand German but may have been deterred by Mendel’s mathematical analysis having once said, according to Galton [2009], “Mathematics in biology is like a scalpel in a carpenter’s shop - there was no use for it.”
where $\rightarrow$ means simple copy  
$\Rightarrow$ means duplication in meiosis  
$\sim$ means mutation in copy  

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Parental Genotype</th>
<th>Resulting Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>aabb</td>
<td>aabb</td>
<td>aabb</td>
</tr>
<tr>
<td>~\sim\sim Aabb</td>
<td>Aabb</td>
<td>AAbb</td>
</tr>
<tr>
<td>~\sim\sim aaBb</td>
<td>aaBb</td>
<td>aaBB</td>
</tr>
</tbody>
</table>

Table 2.1: Viable genotypes in the Dobzhansky-Müller model

It is useful to note that this model does not depend on the principle of dominant and recessive genes, $a$ and $A$ (like $b$ and $B$) are simply required to be different alleles with neither designated as dominant. It should therefore be possible to simulate this process without modelling dominance and recession.

The only reference found in the literature to modelling of the Dobzhansky-Müller mechanism is in Gavrilets [1997], but this was a mathematical analysis rather than a simulation. In that instance the adaptation is designated as $AAbb$ evolving into $aabb$ & $AABB$, which while not the conventional notation nevertheless describes the standard process.

2.3.3 Hybrid speciation

Jiggins [2006], Kronforst et al. [2013], Smith and Kronforst [2013] discuss genomic evidence for the incidence of hybrid speciation which is a form of sympatric speciation occurring when two genetically similar populations cross to form a hybrid which becomes genetically isolated from both parent populations. This is to be distinguished from sympatric bifurcation where a new species develops from within an existing population. Several authors, Brower [2011], Hill et al. [2013], Kronforst [2008], Kronforst et al. [2013], Mavarez et al. [2006], Smith and Kronforst [2013], describe this phenomenon in populations of butterflies of the genus *Heliconius* which exist in Central and South America. In some of the speciation events observed, polyploidy caused the isolation of the hybrids, but in other cases, the hybrids retained the same number of chromosomes as the parent group, indicating that polyploidy had not occurred. This latter type of hybrid speciation seems to be a more interesting research topic than the more obvious mechanism of polyploidy.

Kronforst et al. [2013] identify two key aspects of the hybridization in *Heliconius* butterflies. They are: assortative mating (sexual selection), and mimicry of poisonous species. After random mating, assortative mating seems to be the initial driver for isolating a new population because mating choices seem to be strongly influenced by the colour patterns on the butterflies’ wings. Kronforst et al. [2013] say: "Our analyses reveal that initial divergence is restricted
to a small fraction of the genome, largely clustered around known wing-pattern genes." This initial change is reinforced by the survival of individuals with wing patterns which mimic poisonous species. The survival of non-poisonous species is enhanced because predators tend to avoid eating those individuals with certain wing patterns. Kronforst et al. [2013] suggest that mimicry is reinforced because predators memorise only a relatively small range of ‘dangerous’ wing patterns which results in a predominance of certain patterns.

The steps in the process of hybrid speciation in Heliconius are described by Kronforst et al. [2013] as follows. Firstly, a random cross between dissimilar individuals results in the development of a new wing pattern. Secondly, if this wing pattern is similar to a poisonous species predators avoid eating it and the type persists. Then, assortative mating reinforces the isolation of the group. During this process, the populations are probably not truly separate species as defined by reproductive isolation but only sexual selection is maintaining the separation. Finally random mutations may result in differences in the phenotype which provide insurmountable barriers to successful interbreeding. At this point the group can be defined as a truly separate species.

Kronforst et al. recount that individuals from separate groups can be successfully crossed in the laboratory despite the fact that they would avoid mating in the wild. This is similar to the apparent separation of the flowers Mimulus lewisii and Mimulus cardinalis which are isolated by attracting different pollinators but can be crossed artificially (Charlesworth and Charlesworth [2003]).

Abbott et al. [2013] provide a comprehensive review of hybridization and speciation. In particular they comment, “Hybridization may contribute to speciation through the formation of new hybrid taxa, whereas introgression of a few loci may promote adaptive divergence and so facilitate speciation.”. They go on to write, “It may provide the raw material for adaptive divergence or initiate new hybrid populations, potentially leading to speciation.”.

This process should be distinguished from the Dobzhansky-Müller mechanism (see 2.3.2) where a two-stage change in genetics is required to produce a species barrier. In the hybrid speciation process, described by Kronforst et al., an initial divergence is maintained by assortative mating, then reinforced by selective predation until truly irreversible phylogenetic barriers develop. This appears to require changes at several different levels: an initial genetic mutation, assortative mating by the organism itself and predation by an external agent. It is not simply a genetic process.

### 2.3.4 The influence of time and population size

Some research seems to suggest that long time scales and small populations may be effective at fixing genetic changes so that they become species separations. These effects are worthy of more investigation.
2.3.4.1 Time and the ‘Museum Model’ of the tropics

A key requirement for the diversification of species, whether sympatric or not, appears to be time. This is illustrated by an interesting theory about the observed high diversity of species in the tropics when compared with that in higher latitudes. In this respect the tropics has been described by Gaston and Blackburn [1996] as “a cradle of diversity, a museum of diversity, or some combination of the two.” That is to say, that species are somehow generated in the tropics or preserved there. Galton and Blackburn conclude that the museum model is the more likely and they attribute the greater diversity of species in the tropics to the longer time over which the eco-system has existed there. In higher latitudes the eco-systems have more frequently been degraded by periodic glaciations which effectively reset the regional ecology.

The validity of this concept is still disputed but Condamine et al. [2012] support the idea and say “Amazonia probably allowed the persistence of old lineages and contributed to the steady accumulation of diversity over time ...”, whereas Espeland and Murienne [2011] show that New Caledonia exhibits great diversity despite having been submerged. They say, “In opposition to the museum model, our results provide additional evidence that original New Caledonian biodiversity was wiped out during the episode of submersion, providing an open and empty space facilitating evolutionary radiations.”.

These observations seem to indicate that, in order to achieve speciation, a simulation should encompass a long time scale covering many generation. To achieve this in a manageable research project the model will need to simulate each generation as quickly as possible.

2.3.4.2 Genetic drift in small populations

Lynch and Gabriel [1990] suggest that mutations may be more likely to persist in smaller populations than in larger ones. They write, “In small populations random genetic drift will progressively overpower selection making it easier to fix future mutations.”. They specifically refer to deleterious mutations but this may also mean that speciation could become fixed more easily if a change initially appears in a small population.

These findings suggest that the overall size of the population being modelled may have an influence on the success of speciation. Later simulations confirmed that a uniform initial population often overwhelmed a small colony of mutant individuals.

2.3.5 Summary of mechanisms for speciation

In the Dobzhansky-Müller Model and in hybrid speciation the species barrier is considered to be finally closed by some insurmountable difference between the groups. This could be characterised as a phenotype difference in one of the forms described with reference to Species Boundaries (2.1.2, above).
In summary, assortative mating, phenotype difference, hybrid speciation, time scale and population size are useful ideas which can be tested in the current research to investigate how sympatric speciation might occur without invoking polyploidy.

The next section looks at the selection of a suitable modelling method.
Chapter 3

Literature Review: II - Genetic Modelling

3.1 Modelling objective

The review of biological literature in Chapter 2 indicates a lack of compelling evidence for the natural occurrence of sympatric speciation, without a polyploid copying error, but this does not necessarily mean that it never occurs. It rather leaves an interesting opportunity to investigate the mechanisms of genetics to see if sympatric speciation could occur. In particular, it would be useful to follow up the suggestion by Coyle and Orr [2004], quoted above, that ecological factors might initiate a separation which is later reinforced by reproductive isolation; and the idea from Gavrilets [2003] that an assortative mating mechanism must be present. It is therefore prudent to review the scope of the current project to clarify which of these mechanisms to investigate.

With this in mind, the research should include assortative mating because this may assist speciation but exclude polyploidy which can be seen as a simple, and uninteresting, short-cut to sympatric speciation.

In order to investigate whether sympatric speciation can occur in nature it is necessary to model the natural genetic processes of reproduction as closely as possible. A particular type of model is therefore required to investigate this possibility and more generalised evolutionary algorithms will not suffice if they do not emulate the mechanisms of natural genetics.

There are three key requirements for investigating sympatric speciation:

- the model must ideally represent natural reproductive processes;
- the simulations must create circumstances in which sympatric speciation can occur;
- a criterion must clearly indicate that speciation has occurred.

This second part of the Literature Review looks at existing methods of genetic modelling in general and then concentrates on the history and development of one particularly appropriate form, “The Penna Model”.

36
3.2 Possible modelling techniques

First it is necessary to define the key characteristics of the most appropriate model method for the current research. The required method should emulate the reproductive mechanisms of meiosis but it does not need to become embroiled in the complexities of cell biology. It should allow for some interactions between organisms and between the organisms and the environment. Because it will be used to investigate the logical possibilities of the basic mechanisms of gene exchange and inheritance the model does not need to represent a specific organism or group of species. However, in order to achieve the project objectives the modelling method must be capable of representing multiple stable species at the same time.

3.2.1 Off-the-shelf models

Hoban et al. [2012] reviewed a wide range of software currently available for the analysis and modelling of genetics and evolution. They point out that “Simulations have traditionally been used in population genetics by a fairly small community with programming expertise, but the recent availability of dozens of sophisticated, customizable software packages for simulation now makes simulation an accessible option for researchers in many fields.” However, for the current research project it seems essential to maintain rigorous control of the assumptions and mechanisms employed in the model. The use of an existing software package would probably not provide, with sufficient clarity, the details of the genetic mechanics used. Such a model might also limit the scope of the user to alter key characteristics of the modelling process. For this reason the current research relies on original source code to achieve the required modelling, as described later in this thesis.

Many academic modelling authors publish open-source code which would allow the operation of their model to be understood and possibly adapted to suit the current project. However, time and effort would be required to understand the existing model and to modify the code and test the original functionality and the changes. It therefore seemed easier to undertake the modelling from first principles with a purpose-built model. This approach also seems to accentuate the computing element of the project which sometimes seems lost in the technicalities of what is being modelled.

The two obvious disadvantages of developing original software for the model are: (i) the time taken to develop the model; and (ii) the risk of errors which may go undetected in the ‘original’ code. Against (i), the time-scale, it can be said that learning to use and to trust a proprietary package could be very time consuming and might reveal that the required modelling mechanisms are unavailable in the selected software package. Against (ii), the risk of errors, one must set the possibility of errors in an ‘off-the-shelf’ model in which there may be no possibility of checking the algorithms used, whereas extensive testing of one’s own code can reduce, if not completely
eliminate, the possibility of coding errors. On balance the author of this thesis has chosen to use original code to build a version of an established model which can then be tailored to serve the exact needs of the research project and be tested as extensively as possible.

It also seems a more valid form of research to develop one’s own simulation environment rather than to rely on external sources. As a computing researcher one strives to be part of the “... small community with programming expertise ...” as mentioned by Hoban et al. [2012]

3.2.2 Various modelling methods

Other literature describes models which are not particularly biologically based and therefore seem less promising for the current requirements. Many researchers use methods which could be called evolutionary algorithms but these do not meet the present research requirement to emulate the logic of natural biological reproduction.

For example, Ashlock and von Koningslow [2008] achieve a ring-species using the evolution of a simulated robot which does not seem to truly represent natural, biological, genetic processes as closely as required. Diekmann and Doebeli [1999] propose a model with only limited biological reality. Other research has developed artificial life models such as Ray [1994] which reproduces an evolutionary environment but operating significantly differently from natural biological processes. Papers such as Back et al. [1997] offer interesting background on the history of evolutionary modelling but are not specific enough about how genetic inheritance might be modelled. Jeltsch et al. [2008] describe modelling the spread of plants but do not deal with the issues at a low enough, genetic level. Similarly, Chu [2008] provides for the inheritance of characteristics by passing on parameter values to the next generation. In nature these parameters would be the output of inherited genes but the inheritance of the genes themselves has not been modelled.

In their short 1999 paper Diekmann and Doebeli claim that sympatric speciation is achievable in a digital model of individuals. They use haploid as well as diploid models and employ ‘marker’ genes which interact with the environment to influence mating and reproduction rates. Using these techniques they report the division of the population into separate groups which they describe as species but there appears to be a number of deficiencies in this method.

Firstly, no reasoning is given about whether the separated groups are species, by any definition, and there is no mention of the estimated inter-fertility of the groups. Secondly, the recombination of genes in the diploid model during reproduction is very different from natural processes. Genes are selected at random from either chromosome which is inconsistent with the natural crossover mechanism: to achieve this degree of gene-mixing an unrealistically large number of crossovers would be required. Thirdly, the use of interactions between genes and the characteristics of the environment brings into question the concept of a sympatric region. Lastly, by
assigning differing environmental characteristics to various parts of the region it is questionable whether the region remains truly sympatric.

These shortcomings highlight two important issues for the current project. Firstly, the simulation of ‘unnatural’ crossover mechanisms raises the question of whether the simulations are relevant to the natural world. As mentioned in the Introduction (Section 1.1.2 on page 17) simulations of speciation are only valid as a logical model of the natural world if they closely follow natural processes; otherwise they are simply modelling curiosities. Secondly, the issue of what is truly sympatric remains a difficult semantic question. It could be argued that different life-experiences of two adjacent organisms is an allopatric scenario but this argument would make a sympatric scenario almost impossible to achieve; alternatively an environment across which conditions vary gradually, with no defined barrier, could be considered sympatric which would give ‘sympatric’ a very wide definition. For the purposes of the current project a strict definition of ‘sympatric’ has been used: it is assumed that the same environmental conditions must apply to the whole of an area for it to be considered ‘sympatric’.

In their recent paper Elfwing and Doya [2014] describe a modelling investigation of the emergence of mating strategies within an inter-breeding population of a single species. However, Elfwing and Doya seem to have modelled the phenotype having simulated the behaviour of physical robotic entities which had previously been built and used in smaller numbers. Similarly, Kittas [2010] does not attempt to model the genetic processes. He describes his model as “... a phenotype (in contrast to genotype) centred model and does not focus on a specific genetic procedure that affects the ageing process.”. These types of model do not match the project objective of investigating the genetic processes of speciation.

Balloux [2004], Broom et al. [2002], Orr [2000], as well as others, deal with aspects of evolution by mathematical analysis which is a valid approach but does not suit the aims of the present project, which aims to use a computer simulation rather than mathematical analysis.

3.3 The Penna model

This model, first described by Penna [1995], represents genes as binary digits held on a number of chromosomes, each being a binary number. Mechanisms of meiosis including crossover and some copying errors are allowed for. The Penna type of modelling therefore seems a useful technique which stays reasonably close to biological genetics and uses a digital simulation of genes. It could therefore be developed and modified for the current research.

A fuller description of the workings of the model and how it was implemented and modified for the current project are given in Chapter 4, but some key characteristics are outlined below.
3.3.1 Genes and bits

The bits on the chromosome are taken to represent whole genes, and in some implementations only the genes of interest to the researcher (De Oliveira et al. [2004]). Therefore a relatively small number of bits can represent a large number of base-pairs on the DNA of a real chromosome; the bits should not be taken to represent base-pairs themselves. By their nature the bits can only exist in two states, 0 or 1, and this necessarily limits the genes to two alleles, with 0 representing a ‘wild’ form of the gene and 1 a ‘mutant’ form. However, groups of bits taken together could, to some extent, represent higher numbers of alleles (4, 8, 16 etc.), but crossovers within such a group of bits would de-link the groupings and would complicate the interpretation of multiple alleles. This interpretation does not appear ever to have been reported.

3.3.2 Binary mutations

The only meiotic copying errors represented in the Penna model are changes of a binary digit. These errors are applied probabilistically and generally only change in one direction: from 0 to 1.

The model does not allow for other forms of copying error which can occur during meiosis such as the inclusion or exclusion of segments of the chromosome, often referred to as introns and exons, neither does the model allow polyploidy to occur.

3.3.3 The ageing process

An ‘ageing’ process is a key characteristic of the original Penna model and could be considered its principle distinctive feature. At ‘birth’ a pre-determined group of genes is examined for mutations in both chromosomes, which indicate a fatal disease. At every time-step additional loci become ‘critical’ and are tested for double mutations which simulates the progressive expression of more genes with age which may result in fatal disease. This model has successfully simulated the onset of age-related disease in populations [Stauffer, 2007].

Where the whole chromosome is subject to the ageing process it has been shown by Waga et al. [2007] that, when the probability of crossover is greater than 0.4, the evolutionary tendency is for the ‘cleansing’ of all genetic mutations from the genome, despite the fact that changes from mutated to unmutated alleles (1 to 0) is not permitted. This results from the progressive inclusion of more and more unmutated alleles (0) at the ends of the chromosomes, which is favourable because an unmutated allele on a gamete will partner successfully with any allele (1 or 0). This process relies on crossovers to join more and more unmutated alleles to the ends of the chromosome and compensate for any additional mutations. Consequently, over time, the whole chromosome can become unmutated.
3.3.4 Chromosomes and ploidy

In its original form the Penna model used a single (haploid) chromosome of 32 genes all of which were subject to the ageing process. Later versions used a diploid form and some used multiple sets of chromosomes.

The model can be adapted to represent asexual reproduction where the offspring have only one parent; in this case some form of mutation is required to produce any evolutionary change as no crossover occurs. In sexual reproduction the chromosome of the offspring is generated by mixing the genomes of two parents using a crossover process based on meiosis. This sexual reproduction can be applied to both haploid and diploid genomes.

Sexual reproduction and diploidy seem to be confused by some authors. De Oliveira et al. [2004] seems to imply that a diploid somatic form is required for sexual reproduction which is not the case either in the context of the model or in natural biology\(^1\).

The distinction between sexual and asexual reproduction is not directly related to the sex of the individual organisms. In some implementations of the model [De Oliveira et al., 2004] individuals are somewhat arbitrarily designated ‘male’ or ‘female’ in the context of limiting the selection of mates. This is not the same as modelling the sex-related chromosomes, X and Y in mammals or Z and W in birds, which would require special treatment of these sex chromosomes so that the pairings YY and WW are never implemented. The mechanisms for handling sex-chromosomes are not relevant to the current discussion because they do not seem to have a direct bearing on the process of speciation. They have therefore not been implemented in the current research model.

3.3.5 Population control

In some early versions of the Penna model the population was controlled by a Verhulst function:

\[
\frac{dN}{dt} = r.N\left(1 - \frac{N}{N_{\text{max}}}\right)
\]

where:
- \(N\) is the current population
- \(t\) is time
- \(r\) is the underlying growth rate
- \(N_{\text{max}}\) is a preset maximum population

\(^1\)In nature mosses are haploid in their somatic cells and diploid in their ‘seed’ (spores), whereas mammals are diploid in the somatic cells and haploid in their ‘seed’ (gametes), but both mosses and mammals reproduce sexually.
Applying this algorithm has the effect of slowing growth as the total population approaches the maximum value. It was implemented in the Penna model as a death rate of $rN/N_{\text{max}}$ randomly applied to the whole population at each time-step.

Stauffer [2007] describes the use of the Penna model to simulate the onset of death by ageing and compares the results with actual demographic data from Sweden and Germany. He refers to the original haploid form of the model but also to a diploid sexual version with chromosome crossover during reproduction. In Stauffer's versions of the model the population is limited by a Verhulst function as Penna used in his original, but Stauffer also refers to work by Waga et al. [2007] which investigates the effects of various probabilities of crossover. In Waga's model the population is limited to a fixed lattice without a Verhulst function. It is not clear when the lattice structure was first introduced into the model.

Later versions of the model, for example Waga et al. [2007], use a rectangular 'lattice' of fixed dimensions with only one organism permitted in each cell of the lattice. This limits the population to the size of the lattice. This form of population control is used in the current research.

### 3.3.6 Monte Carlo simulation

Some authors including Waga et al. [2007], and Penna and Stauffer [1995] describe the Penna model as a *Monte Carlo* simulation. This term is more often used to describe calculations which are repeated many times, using randomly selected inputs to give a range of possible outcomes.

Landau and Binder [2009] describe a Monte Carlo simulation as one which "... depends on a sequence of random numbers which is generated during the simulation. With a second, different sequence of random numbers the simulation will not give identical results but will yield values which agree with those obtained from the first sequence to within some 'statistical error'". The key characteristic of the Monte Carlo method is the collation and analysis of a large set of results, from typically thousands of simulations, to give a distribution of possible outcomes. The Penna model traces the life of each organism using randomised events including inheritance, death etc., and develops a changing population.

This is not the same as repeating one calculation with a controlled range of input parameter values. The simulation would only become a true *Monte Carlo simulation* if the results of numerous simulation runs were collated. This is not generally the method described for using the Penna model and it therefore is not properly described as a Monte Carlo simulation. Waga et al. describe each time step in the Penna model as a "*Monte Carlo step*" which does not conform to the conventional usage outlined above. However, if the model is rerun several times using different randomisation seeds then the process becomes similar to a Monte Carlo simulation, but this was not what Waga et al. report.
In the current research each simulation is generally rerun 10 or 20 times and the results are averaged. Details of the methods used are reported in Chapter 5.

3.4 Origin and development of the Penna model

This section gives a brief history of the origins, development and various uses of the simulation method known as The Penna model.

3.4.1 The original form of the Penna model

In 1995 Penna published details of a newly devised “bit-string model of biological ageing” in a paper which is surprisingly short considering its radical impact. More detail of the model was given in a separate paper by Penna and Stauffer [1995]. The model they described simulates “senescence”, the deterioration of organisms with ageing.

Penna [1995] quotes biological observations to support the results of his modelling. He says that the model shows that “The average age at death decreases as soon as the age at reproduction increases; consequently the total population decreases.” and that “This behaviour has been found in Physella virgata virgata snail populations.”. In this scenario, the reproductive age increases (in the model as a parameter, but in nature because of the presence of Orconectes virilis crayfish) and a corresponding increase in the age at death is observed in nature and in the model. However, the overall population decreases because the effect of the delay in reproduction outweighs the increase in life-span.

Penna and Stauffer [1995] give details of the computer model referred to in Penna [1995]. The descriptions of the model differ in one respect in that: Penna [1995] refers to the generation of ‘one baby’ whereas Penna and Stauffer [1995] refer to one child with probability \( m \). It is not clear from the (unannotated) code given in the paper whether the births are probability driven or not.

The code was written in FORTRAN for multi-threading on a 128-node array of processors and uses a 32-bit binary number to represent 32 critical genes in the genome of each organism. The paper refers to the high computational efficiency of the code and shows that populations of up to 5 billion can be handled but one wonders if such an extremely large population is necessary to demonstrate the phenomena observed. The research for the current project has generally used populations of only a few thousand and found worthwhile results.

One feature of the original model which seems to differ from subsequent implementations is allowing genes to switch from the ‘mutated’ (represented as 1) to the ‘wild’ (represented as 0) allele, as well as vice-versa. In most later versions of the model a ‘wild’ unmutated gene may...
change during the reproductive process to its ‘mutant’ allele but cannot revert to the wild form except by substitution in a recombination process.  

Penna and Staufer [1995] mention that the number of mutations may be increased or decreased during inheritance but elsewhere refer to the case when “... all new mutations are negative...”. This implies that in some cases they set the model to allow only the addition of mutations and never allow reversion to the wild form but in other cases allowed mutations to be reset to the ‘wild’ (0) allele. The one-way introduction of mutations appears to have been generally used in later versions of the model, including the work of Waga et al. [2007].

Using the ‘negative only’ mutation mechanism Penna and Staufer compare their results with those of Lynch and Gabriel [1990] who maintain that the progressive accumulation of mutations will result in the extinction of the population in what they call ‘mutation melt-down’. Penna and Staufer suggest that in their models sufficient children without a damaging mutation can survive to maintain the population. The results from Lynch and Gabriel [1990] are based on a mathematical analysis of the accumulation of mutations rather than a simulation.

Penna and Staufer show outputs in terms of “survival rate”, $S$:

$$ S = \frac{N_t}{N_{t+1}} \quad (3.2) $$

where $N_t$ and $N_{t+1}$ are the population at successive time steps. This is a useful measure of population change.

In the original form of the model there is only one chromosome which is inherited from a single parent with random changes to a limited number of genes in the reproductive process to represent mutations. Penna refers to this as asexual reproduction and notes that a sexual reproduction process could easily be added to the model, as was done in later versions, e.g. by Waga et al. [2007].

3.4.2 Subsequent developments and uses of the Penna model

Since its inception in 1995 various developments and uses of the Penna model have been reported in the literature. The more significant of these are summarised below.

3.4.2.1 Speciation modelling

Speciation does not appear to have been previously investigated using the Penna model but, in their paper, Waga et al. [2007] describe an investigation of the relative impacts of some

---

2 There appears to be an error in the original description of the ageing process which says: “... if an individual has the $i$th bit in genome set to one, it will suffer a deleterious mutation at age $i$.” This should read: “... if an individual has the $i$th bit in genome set to one, it will suffer a disease at age $i$,” because the mutation will have existed from ‘birth’ but its detrimental effect is not manifested until age $i$. 

---
of the key parameters in the reproductive process, specifically the probability of crossover (also known as recombination rate) and the probability of random mutation during gamete production. They also refer to speciation but use the Hamming distance between genomes as a proxy for relatedness. While this seems a reasonable device it can be shown that, in the Penna model, Hamming distance is a misleading indicator of speciation; this is discussed in Section 4.5.1.1 and analysed in Appendix B.1.4. These observations leave an opportunity for the current research to clarify the identification of speciation in this type of model.

Another variant of the model was used by Mackiewicz et al. [2010] to investigate the influence of recombination probability on gene coding density. Their conclusions relate to the “purification” process described by Waga et al. [2007]. This process results from the universal compatibility, in Penna’s model, of unmutated sections of gametes. Where these arise by crossover, they are likely to be more successfully reproduced than gametes with mutations present and with a single crossover this process is most likely to occur at the ends of the chromosomes. Consequently mutations are progressively excluded starting from the ends of the chromosomes3. This provides a useful insight into a common phenomenon in this type of chromosome modelling which has been observed in several of the simulations described Chapter 5 of this thesis.

3.4.2.2 Using ageing processes

The following examples show how the Penna model and its derivatives have been used to investigate ageing and related topics. These processes do not directly influence the development of a model for speciation but they illustrate the scope of model and its correlation with observations in the natural world.

In a short 1996 paper “Why trees live longer?” DeMenezes et al. [1996] use the ageing mechanism in the model to consider the success which trees have in achieving great age. They characterise trees as iteroparous (i.e. with multiple reproductive episodes) and assume that reproductive fertility (measured as offspring per breeding season) increases with age either linearly or parabolically. In both cases they find that increasing fertility with age results in a greater average life-span of individuals. The authors seem to miss the obvious mechanism relating age to fertility in trees, namely that perennial flowering plants, such as trees, increase their fertility with age because as they grow larger they have more reproductive organs (flowers) on them. By contrast, animals generally maintain only one reproductive organ as they grow, although some, fish for example, might produce more eggs as they grow larger. In this case the use of the ageing mechanism in the model seems to be misleading as it does not account for the structural growth of flowering bodies on the trees a point which is not missed by Vaupel et al. [2004].

3This progressive purification is illustrated in Figure 5.2 from a current simulation.
4The authors have unaccountably used an unnecessary question mark in their title.
Vaupel et al. [2004] discuss in detail the mechanisms of ageing and fertility and show that nature provides many examples of fertility increasing after reproductive age is reached. This they call negative senescence which contradicts the assumption made in the Penna model that ageing is associated with the progressive expression of life-threatening mutations as age increases. It had been assumed that evolutionary selection will tend to protect species from diseases in the period prior to reproduction but that beyond that age mutation-related diseases tend to reduce the subsequent life-span. However, Vaupel et al. [2004] demonstrate that if fertility is high enough it will ensure the propagation of genes which protect life after the onset of reproduction.

Bernardes and Stauffer [1996] extended the model to allow for more than one disease per year. Stauffer et al. [1996] compared results from a Penna-type model with those from Redfield [1994] and concluded that sexual reproduction is shown to be more evolutionarily successful than asexual. This probably arises because the Y male chromosome has no matched partner, unlike the XX pair in females, and therefore has fewer sources of unmutated genes and a consequent greater probability of mutation damage. The Penna-type modelling referred to was that reported by Bernardes and Stauffer [1996]. Redfield's mathematical modelling investigated the effectiveness of sexual reproduction at counteracting mutations which arise much more frequently in the gametes of males than in females.

Penna et al. [1999] returned to the ‘bit-string model’ to simulate the effects of age-related disease on populations. They concluded that “the bit-string model has shown [sic] to be a good tool for studies of age-structured population dynamics, in both asexual and sexual versions.” They claim to have reproduced, using their model, the age-rated population profile of people in Germany. It is worth noting at this point, that Penna et al. [1999] are clearly assuming that their model is dealing with one species, in the examples quoted they also refer to modelling the hunting of wolves and over-fishing of cod, in both cases one species is being modelled with man as the external predator. The question of the inadequacy of the basic Penna model for representing a multi-species scenario is discussed in detail later in Chapter 4, the development of the model.

3.4.2.3 Using Verhulst in the Penna model

Different users of the Penna model have employed different methods of limiting the overall population. Initially a Verhulst function was used but in later version of the model the population space was limited to a fixed ‘lattice’ as used by Waga et al. [2007].

Dabkowski et al. [2000] discuss the use of the Verhulst factor \( V = 1 - \frac{N(t)}{N_{\max}} \), in a version of the Penna model, to control population to a maximum size, \( N_{\max} \). Dabkowski et al. investigated the development of static ‘families’ which were identified by residual groups of genes which the authors compared with mitochondrial DNA in humans.
Penn and De Oliveira [1995] also analyse the influence of the Verhulst factor, mutation rate and reproductive age, as used in the Penna model, and derive a mathematical expression for the stable population size for semelparous (single-breeding) species. They claim that this gives results which correspond with field observations for Pacific salmon and mayflies.

Piñol and Banzon [2011] discuss the effects of using a Verhulst type of growth limitation in the Penna model as against allowing old-age and space to limit the population. Their objective was “to find stable populations without having to impose a Verhulst factor.”. By using age-dependent death and limiting reproduction to once per lifetime, they were able to achieve a stable population which neither collapsed nor grew without limit, without using the Verhulst factor. In the current project, the population is constrained only by space and generally expands to fill the available space unless depleted by environmental factors, in which case the population will die out. When using space to limit the population, the population will either die away or will stabilise at the maximum available capacity less any immediate deaths. This happens because to stabilise at an intermediate population size would require a perfect balance between births and deaths over a sustained period which is very unlikely in a randomised modelling process.

One variant of the model was used by Cebrat and Pekalski [2004] to investigate adaptation to environments of varying hostility in which the differences between the number of mutations in an organism and those in a genotype ‘ideally’ suited to the environment determined the organism’s survival. They report the calculation of the ‘adaptation’ factor, $a_j$, for organism $j$ as:

$$a_j = \frac{1}{32}(1 - \sum_{\alpha=1}^{32} \text{XOR}(f_{\alpha j}, F_{\alpha}))$$  \hspace{1cm} (3.3)

where $f_{\alpha j}$ is at locus $\alpha$ on the chromosome of organism $j$ which may be 0 or 1, $F_{\alpha}$ is the corresponding locus in the ‘ideal’ genotype, and XOR() is the exclusive OR function which counts differences between the genotypes. The probability of survival is then calculated as: $p_j = \exp(-sA_j/a_j)$, where $s$ is a factor representing selection pressure and $A_j$ is the age of organism $j$. There would seem to be an error here if one assumes that $a_j$ is intended to be in the range 0 to 1. The maximum possible number of mutations is 32, being the length of the genotype, and therefore the adaptation, $a_j$ would be better calculated as:

$$a_j = \frac{1}{32}(32 - \sum_{\alpha=1}^{32} \text{XOR}(f_{\alpha j}, F_{\alpha}))$$  \hspace{1cm} (3.4)

which would return 1.0 if all 32 genes matched and 0 if none matched. This technique is similar to, but not the same as, the adaptation algorithm used in the current research model described...
in Section 4.4. There, mutations at ‘beneficial’ loci afford an organism greater tolerance of a hostile environment, but the probability is not based on an age-dependent function as Cebrat & Pekalski’s used.

3.4.2.4 Other uses of the Penna model

Thoms et al. [1995] used the original haploid form of the model but extended the mechanisms by introducing non-heritable somatic mutations, and related the death-rate to a comparison of the mutations of an individual and the average mutations in the population. There would seem to be a good case for using somatic mutations which implies that some genetic defects occur during the life of an organism and do not only occur during reproduction. Instead of counting all mutations as potentially deleterious they compare the mutations with a pre-set ‘optimal sequence’ of genes which is assumed to represent optimal fitness for the prevailing environment. This seems more realistic than the method used by others which assumes that genes become deleterious with age in the order in which they appear on the chromosome. This might be unimportant until diploid crossovers are introduced which tend to favour the transmission of genes nearer the middle of the chromosomes and changes to genes at the ends.

Thoms et al. also use a limited span of fertility between the ages of 11 and 25 years. Their results show an interesting rising mortality in the early years of life, then a decline followed by a rise at the end of the fertile age. They claim that this is the first report of modelling showing two maxima in mortality and that this may represent antagonistic pleiotropy in which a mutation which is beneficial in youth becomes deleterious in later life. This idea must be contrasted with negative senescence identified by Vaupel et al. [2004] referred to above (3.4.2.2).

Penna et al. [2001] used the model to investigate the impact of over-fishing on lobsters and damselfish populations. Haddad and Penna [2008] reported on using a version of the model to investigate the effects of a change in fertility as might occur if a predator was suddenly removed from a stable eco-system. These uses of the model illustrate the wide variety of applications for which it has been used. DeMenezes et al. [1996] also refer to analytical work by Ito [1996] which relates birth-rate and population density in the model. Puhl et al. [1995] describe extending the use of the model to consider predator-prey interaction.

The breadth of these studies illustrates the flexibility of the Penna model for investigating evolutionary development and shows that in many cases it has been successfully compared with data from the natural world. It does not seem to have been previously used for speciation research.

3.4.3 Summary of the history of the Penna model

In a short paper De Oliveira [1998] and more extensively in De Oliveira et al. [2004], De Oliveira summarised the developments of the Penna model since its original publication in 1995.
This group of authors, which includes Penna himself, also provide some useful clarification of the principles behind the model. For example, the limited range of genes (32) is not intended to represent the whole genome of any organism but only those genes which are critical to the ageing process. Similarly, it becomes clear that the time scale of the model is not intended to be one breeding season per time step, because many offspring may be ‘born’ in a time step indicating that each step could be considered to span one or more breeding seasons. These clarifications, in Stauffer [2007], help to explain the success of the model in its intended application to modelling the ageing process of individuals of a single species.

One feature of the model which is not commented on in the literature is the logic of consecutive accumulation of ‘critical’ gene loci with ageing. In the model, as age progresses, adjacent loci are consecutively included in the ‘critical’ group which may cause disease. The fact that these loci are adjacent is acceptable because one may assume that other ‘less interesting’ genes lie between them and do not need to be represented in the model. However, the fact that the loci are consecutively added to the critical group as age progresses, is less defensible because, in nature, genes for age-related diseases may be located in any order on the chromosome and during crossover they may therefore be separated or combined differently from a consecutive group of genes. This inadequacy could be overcome by bringing the loci into the critical group in a preset, consistent order which would not necessarily be the order in which they appear on the chromosome.

Key input parameters identified by De Oliveira et al. are: the reproductive age range (which may have a maximum as well as minimum); the number of births possible in each cycle; the probability of random mutations in each reproductive process; and the number of mutations which can be tolerated before they become fatal.

De Oliveira et al. describe triploid organisms and say that these can be implemented in the model. These can be infertile in nature because of the unbalanced division of the odd number of chromosomes during the anaphase of meiosis but the ‘edible frog’ Pelophylax esculentus is somatically triploid and has evolved complex hybrid gene exchange with related populations, as described by Christiansen [2009] and discussed in Section 2.2.2. The authors also discuss the implementation of parasexual processes which involve haploid organisms achieving genetic mixing by passing through a diploid phase.

However, with reference to the diploid form of the model they say: “To count the accumulated number of mutations and compare it with threshold $T$, we must distinguish between recessive and dominant mutations. A mutation is counted if two bits set to 1 appear at the same position in both bit strings (inherited from both parents), or if it appears in only one of the bit strings but at a dominant position (locus). The code randomly chooses the dominant positions at the beginning of the simulation; they are the same for all individuals.” This concept of specifying the loci where dominance operates does not conform to the usual biological concept of dominant and recessive genes. The normal model, originating from Mendel (see Suzuki et al. [1981] Chapter
2), requires recessive genes to be present on both chromosomes to be expressed in the phenotype but dominant genes are expressed if only present on one. The designation of some specific loci as the ‘dominant’ ones is an artificial construct. In later forms of the model, e.g. Waga et al. [2007], mutant genes (1) are treated as if they are recessive and only become deleterious when they are present on both chromosomes and the wild forms (0) are treated as dominant. This is the preferable arrangement because it conforms to the standard Mendelian concept of dominance and recession. The sexual version of the model randomly assigns a sex to each individual, but this is not related to chromosomal differences as in biology, and seems, according to De Oliveira et al. [2004], to have no substantive impact on the modelling outcomes. In later versions of the model the sexing of individuals is discontinued.

De Oliveira et al. [2004] describes inserting random mutations after the creation of the gametic chromosomes. In nature, mutation errors in meiosis are more likely to arise during the copying of the chromosomes before crossover which is before gametic chromosomes are formed. The important difference from De Oliveira et al.’s model is that if errors are introduced only in copying the parent chromosome then transmission of these to the gamete may be excluded from some gametes depending on the mutation’s position relative to the crossover. De Oliveira et al.’s method of potentially placing a mutation in every gamete will therefore tend to introduce more mutations than would arise during copying.

De Oliveira et al. [2004] discuss sympatric speciation in a sexual diploid version of the Penna model. They suggest that to achieve this “... the goal is to start with one population and at the end have two populations coexisting with each other in stable equilibrium but without cross mating”. The last phrase without cross mating is the essential criterion which concurs with the definition of speciation derived from other literature, see section 2.1.2.1.

This contrasts with the association by Waga et al. [2007] of Hamming distance with speciation. Waga et al. use a particular gene as an indicator of the potential difference between species. This gene is used, simultaneously, in two ways: firstly, it is used to match mates thereby implementing assortative mating and secondly, it attracts different death rates by invoking different parameters in the Verhulst function. This process results in the separation of the groups, but it could be argued that modifying the Verhulst function is equivalent to allopatric speciation where populations in different areas are subject to different pressures. The use of assortative mating on genes also defining the probability of survival concurs with the suggestion of Gavrilets [2003], noted above, that speciation cannot be achieved without assortative mating on genes linked to disruptive selection which in this case is achieved by linking sexual selection and the Verhulst function to the same gene.

These mechanisms are similar to the development of insurmountable species barriers which are described in section 2.1.2, above. However, it may be important to draw a clearer distinction,
than De Oliveira et al. [2004] do, between genes which are important for assortative mating and those important for insurmountable phenotype differences. The former may be thought of as ‘soft’ barriers imposed by the sexual ‘choice’ of the organisms, whereas the latter as ‘hard’ physiological differences which prevent inter-breeding and enforce the species separation.

De Oliveira et al. [2004] claim (without giving any supporting data) that “the Penna model is by far the most used computational ageing model...” and it is true that there have been over 300 citations of the original paper [Penna, 1995] since it was first published. The principal users of the model seem to be concentrated in three groups: one with Penna himself in Brazil, one with Stauffer [2007] in Germany and another with Cebrat and Pekalski [2004] in Poland. However, there is little reference to its use in the study of speciation.

3.5 Proposed modelling method

From the work of others it seems apparent that the Penna model provides a useful basis on which to undertake the present research and allows the flexibility required to meet the objectives set out at the beginning of this chapter. However, this model has not been used effectively for speciation research and some modifications are likely to be required.

Principal advantages of the model are that it includes:

- genes represented individually,
- diploid chromosomes,
- meiosis including crossover,
- single gene mutation (although other copying errors are not modelled),
- a spatial lattice which limits the physical range of gene exchange,
- the lattice also serves to limit the population.

In order to model speciation, features which need to be improved are:

- assortative mating,
- phenotypic differences between species,
- predation and mimicry,

The implementation and development of the model is described in Chapter 4, Development of the model.
Chapter 4

Development of the *Sympatria* model

Reporting of the current research project is divided into two parts. The first part, in this chapter, describes the computer model and its development. The second part, in Chapter 5, summarises some of the simulations undertaken with the model and their results.

More details of workings of the model are given in Appendix A and some theoretical analysis is included as Appendix B.

4.1 The development process

This chapter is divided into the following sections:

- 4.1: The development process: this introductory section;
- 4.2: The basic Penna Model: the model as described by Penna [1995], Waga et al. [2007] and others;
- 4.3: Implementation of the *Sympatria* Model: a new implementation of the model;
- 4.4: Adding environmental variation: initial additions to the model;
- 4.5: Recognising colonies: how to identify potential species;
- 4.6: Modelling multiple species: extending the scope of the model;
- 4.7: Selection mechanisms: introducing different mechanisms;
- 4.8: The Dobzhansky-Müller mechanism: testing the development of species.
4.1.1 Limitation of modelling

All computer modelling is an approximation to reality and the model used in this research makes many assumptions about the interaction of genes and organisms. For example, it models genes as entities whereas they have complex biochemical structure and genes exist in cells the chemical processes of which are only just being understood. No attempt is made in the current model to show a causal link between genes and the phenotype organism which they create, but assumptions are implicitly made about the nature of the phenotype from its genes. Modern research into epigenetics indicates that, in many organisms, the form of the phenotype is not solely defined by the genotype, but also by environmental and other factors. However, in the context of the current project, it is assumed that some significant features of the phenotype can be deduced from the genotype.

Most importantly, the results from this modelling of genes have not been compared with any data from living organisms. This project is, in essence, a series logical experiments on the possible development of genetic changes during evolutionary processes. It seeks to establish, in principle, the possibility of sympatric speciation occurring. It does not aim to associate these mechanisms with any specific types of organisms, but rather with any diploid, sexually-reproducing species.

The validity of any deductions made from the model depends upon the model being a reasonably true representation of the natural world. Without this validity the model is simply a computational curiosity.

There are numerous records in the literature of the use of the Penna model but very little discussion of its validation against natural processes. For example Stauffer [2007] provides a comprehensive, and generally complimentary, assessment of the Penna model. In their paper on sympatric speciation Waga et al. [2007] use the model with an implicit assumption that it represents nature.

These concerns indicate that development and use of the proposed model must proceed with caution bearing in mind these limitations.

4.1.2 Interpretation of the model

The descriptions of the Penna model by others often use the terms ‘cell’, ‘mate’ and ‘baby’ in describing the model and these authors’ models were used to investigate animal genetics and in some cases human genetics, for example: Biecek and Cebrat [2008]. Following this earlier work, much of the original coding of the present model maintained that terminology and it is used in the names for some data structures.

However, the model is clearly not a model of a cell because most single cellular organisms do not reproduce sexually, neither does the model represent an animal because the ‘organism’ remains
static in the lattice after it is ‘born’. It would therefore seem to be a truer interpretation to consider the model as representing plant genetics because the ‘organisms’ are static and populations only spread by setting seed within a defined range from the ‘mother’ plant. ‘Mating’ is also limited to the locality of the ‘mother’ plant and consequently more like pollination. Therefore in this report, the terms ‘plant’ (or ‘organism’), ‘pollination’ and ‘seed’ have generally been used, but in the coding of the model some of the old terminology persists.

4.2 The basic Penna Model

As outlined at the end of Chapter 3, the model originated by Penna [1995] and developed by others was chosen as the starting point for the current modelling process.

The work of Waga et al. [2007] indicated an interesting possibility of researching the development of genetically separate groups of organisms in a fixed space. This required the development of the model and then the design of appropriate simulations to test for the development of sympatric speciation. The type of model developed by Waga and others after Penna offers a convenient basis on which to investigate sympatric speciation.

The model is based on an idea by Penna [1995] and Penna and Stauffer [1995] which has been developed by Waga et al. [2007]. It is a high-level computer simulation of the interaction of diploid entities (analogous to plants) which occupy a rectangular lattice, typically a few hundred units across. Each position in the lattice can hold only one ‘plant’ and therefore reproduction can become limited by space. There is a universal death rate, applied randomly, which has the effect of creating space for new seedlings.

The diploid chromosomes in each ‘plant’ consist of a number of genes each represented by a binary digit, ‘0’ indicating the wild allele and ‘1’ a mutated allele. Consequently only two alleles are possible at each locus. For computational simplicity a chromosome of 64 genes, represented by a 64-bit number, has been used throughout the current simulations. The lattice may be ‘wrapped’ in either or both dimensions to create cylindrical or toroidal surfaces, so that for pollination and seed-placing, the opposite edges of the lattice are effectively adjacent. The use of cylindrical continuity has the advantage of reducing the number of cells which are needed to represent a large space, thereby speeding up the simulation.

Plants can be pollinated by neighbours within a defined ‘pollination-range’, a square of specified size around the ‘mother’ plant. The resulting seeds are placed randomly in any vacant position within a specified ‘seed-range’. The ranges for pollination and seed placement are input values. Seedlings cannot develop if no vacant position exists within the specified range. A specified number of seeds may be produced by each plant in one time-step.

One of the key attractions of the Penna type of model for evolutionary research is that it allows some spatial limitation on the spread of genes by limiting the range for one plant to
be pollinated by another. That is to say it is not a panmictic model in which genes would be randomly exchanged across the whole population. The spatial limitation of genetic mixing in the Penna model offered a promising tool for investigating sympatric speciation because of the possibility of the formation and consolidation of local colonies of a genotype which might become a species.

4.2.1 Gametes and mutations

The model incorporates mechanisms for generating gametes from copies of parent chromosomes which may include one or more crossovers and copying errors. The number and positions of crossovers and mutations are randomly applied according to probabilities which are input values. The only type of mutation which can arise in the model is in copying individual genes and this may only change a gene from the wild (0) to a mutated allele (1), and not vice-versa. There are no mechanisms for intron or exon mutations in which segments of the chromosome are duplicated or omitted during copying. There is also no mechanism for autopolyploidy when whole chromosomes are erroneously duplicated during meiosis.

4.2.2 Time-steps

The whole lattice is cyclically scanned, in a randomised sequence to avoid directional preferences, and each scanning cycle is considered a time-step. This time-scale defines the ageing of the plants.

4.2.3 Death in the Penna model

Every time-step each plant is also subjected to environmental and genetic tests which determine its survival. Dead plants are then removed from the lattice thereby creating space for new seedlings. There are two independent causes of death, characterised as: genetic and environmental.

Genetic death occurs when the number of double mutations at a predefined group of loci of the chromosome exceeds a defined tolerance. The group of critical loci which are tested for double mutations may increase with age so that an individual may carry a double mutation for a number of time-steps until that locus becomes ‘critical’ and the plant dies. This mechanism is referred to as the ‘ageing process’ and is fundamental to much of the earlier work on the Penna model by others. The number of critical loci for a new zygote and the additional loci per time-step are input values.
Environmental death arises from the random application of a universal death-rate to the whole population in every time-step. This facility was extended later to respond to a varying environment (see Section 4.4) and to allow for predation (see Section 4.7.3).

In some earlier versions of the Penna model a Verhulst function was used to limit the population but in the present model the lattice serves to limit the population.

4.3 Implementation of the *Sympatria* Model

None of the sources found in the literature review give details of the coding of their models and so a completely new implementation of the model was developed for the current research based on the principles described by Penna and others. This was called *Sympatria*.

4.3.1 The development environment

The *Sympatria* model is coded in C++, which the author learned for this purpose. Fuller details of the structure of the model and its object classes are given in Appendix A.3.

During a visit to Poland in 2010, the author discussed the model with Professor Cebrat and his team in Wroclaw, Poland, and learned something of their versions of the model but the *Sympatria* model was written entirely by the author specifically for the current project.

4.3.2 Interfacing with the model

Details of the input to and output from the model are given in Appendix A.3.1.1.

The input was controlled by a text file which allowed all relevant variables to be set to single values or ranges of values; when omitted a default value was assumed. The input files used concatenated text to identify inputs and the meanings of these, and their corresponding mathematical symbols, are shown in Appendix A.1 The symbols listed at the start of this thesis are used in mathematical calculations.

Output from the model was by numerical data in text files and associated graphs, produced using the *gnuplot* package. In some cases the output was given in .csv files which were then post-processed to analyse the results.

All input and output filenames were prefixed with unique codes to facilitate sorting and storage of results. For simulations which gave interesting results the input and output data are reproduced in Appendix C.
4.3.2.1 Colour mapping of genotypes

A technique used by Waga et al. [2007] and others to illustrate the changing genotype distribution in the lattice was colour mapping. This is implemented by specifying 24 loci in the chromosome, in three groups, which map to the RGB 24-bit colour palette. The resulting colours are used to mark each plant on an image of the lattice.

This technique, while giving a general impression of the variation of genotypes across the lattice, cannot fully represent the range of possible genotypes from a 64-bit genome because only a 24-bit sub-set of the genome can be displayed on one map. By allowing the user to specify the loci used to drive the red, green and blue components of the colours it is possible to display maps showing the status of selected interesting parts of the genome, but more subtle differences in genotypes are not evident in such maps.

The mapping also has to deal with possible differences between the diploid chromosomes and some compromises are required here. For example, it was assumed that a genotype with diploid chromosomes A&B is substantially the same as one with B&A, and consequently these may be shown on the image as the same colour. This can be achieved by only showing the colour representation of the chromosome with the larger numerical value.

It later became interesting to distinguish genotypes with one and two mutations at the same loci (see the Dobzhansky-Müller simulations in Section 5.4). To achieve this the colours were coded to show brighter colours for double mutations than for single ones. Details are given in Appendix A.3.1.7.

In the Sympatria model, mapping of the lattice was extended to create other outputs showing variations in compatibility and colonies of compatibility; these are described in Section 4.5.2.1, below.

4.3.3 Randomisation

Many of the algorithms in the model were randomised. For example the position of the crossover in meiosis must be chosen randomly within the number of genes on the chromosome. Other functions required random choices to be biased, such as when mate selection is affected by assortative mating.

Wherever randomisation was required, the compiler’s standard internal random number generator, rand(), was used to provide pseudo-random numbers and its seed was one of the input values. This ensured that a given input file would always generate the same modelling outcome.

Where simulations were run using different input assumptions the same number was used to seed the random number function for each simulation so as to maintain the closest possible similarity between simulations. In general each simulation was rerun several times, sometimes up to 40 times in order to observe any trends in behaviour.
When several sets of these reruns were required with different input assumptions, a random sequence of random numbers was recorded before the first run so that the same sequence of numbers could be used to seed the random number function at the start of every set of runs.

4.3.4 Speed of the model

In coding the model great care was taken to reduce the computational processing time. Details of some of the techniques used are given in Appendix A.3.5. In its final form the model could process about 500000 cell-time-steps per second.

4.3.5 Testing the model

At every stage of its development the model was extensively tested to ensure that the internal mechanisms were performing as intended. Once the proper operation of the code had been established, the additional output code was removed or ‘commented out’ so as not to complicate future outputs or slow the modelling process.

4.3.6 Modification to the basic model

The subsequent parts of this chapter describe the enhancements made to the original Penna-form of model in the course of the current research project. They are described in the order in which they were implemented. The reasoning behind the changes is given in each case as well as a brief description of how the mechanisms were implemented in the model.

4.4 Adding environmental variation

The possibility of different species arising in different places was investigated by allowing for a varying ‘environment’ across the lattice.

This was implemented by introducing an ‘environmental value’ \( V \) which could vary across the lattice and which might be analogous to temperature, humidity or the availability of certain food, for example. The model could be set up with \( V \) varying across the lattice according to various functions, e.g. linearly or in steps.

The general environmental death rate \( (D_0) \) was increases by \( \Delta D_V \) for every unit of \( V \) when \( V \) was outside a ‘comfortable’ range \( V_L \) to \( V_H \), so that:

\[
D_v(V) = D_0 + \Delta D_v (V - V_H) \quad \text{where } V \geq V_H \tag{4.1}
\]
\[
D_v(V) = D_0 \quad \text{where } V_L < V < V_H \tag{4.2}
\]
\[
D_v(V) = D_0 + \Delta D_v (V_L - V) \quad \text{where } V \leq V_L \tag{4.3}
\]
This is illustrated in Figure 4.1.

![Diagram showing environmental death rate (D) as a function of environmental value (V).](image)

**Figure 4.1**: Environmental death rate \( (D) \) as a function of environmental value \( (V) \) used to change the death rate in different environments.

This disjointed form of function was used instead of, for example, a smooth parabolic function, in an effort to induce an abrupt change in the response of the system.

A plant could acquire tolerance to extremes of \( V \) if it had mutations in specific loci, defined by the masks \( L_L \) and \( L_H \). The range, \( V_L \) to \( V_H \), in which there is no enhanced death rate, was increased in proportion to the number of mutations at the appropriate loci, as follows:

\[
V_H = V_{H0} + \Delta V_m \beta(L_H \cap (H_0 \cup H_1)) \tag{4.4}
\]

\[
V_L = V_{L0} - \Delta V_m \beta(L_L \cap (H_0 \cup H_1)) \tag{4.5}
\]

where:

- \( H_0 \) and \( H_1 \) are the haplotype chromosomes of the plant;
- \( \beta(h) \) is the count of bits set in \( h \);
- \( \Delta V_m \) is the increased tolerance to \( V \) per beneficial mutation.

Simulations using this technique are described in Section 5.2.

### 4.4.0.1 Environmental variation and allopatry

The mechanism of the varying environment showed, as expected, that different genotypes developed in response to different environmental conditions, see the results in Section 5.2. However, the development of different genotypes in different environmental conditions does not seem to appropriate for the demonstration of sympatric speciation.
It can be argued that imposing a range of environments on the population tends to create an allopatric or parapatric scenario which is not truly sympatric. Therefore, as mentioned in Section 3.2.2, for the purposes of the current project, a strict definition of ‘sympatric’ has been used: it is assumed that the same environmental conditions must apply to the whole of an area for it to be considered ‘sympatric’.

4.5 Recognising colonies

In using the model it became apparent that a systematic way was required for recognising the development of potentially separate species. This section addresses how the recognition of such ‘colonies’ was attempted.

4.5.1 Data complexity

The model generates large volumes of data. The data can also be very varied and complex because of the probabilistic nature of the modelling process. Consequently it is not easy to identify groups of related genotypes which might vary to a greater or lesser degree.

Earlier uses of the model, reproducing the work of Waga et al. [2007], produced patches on the coloured gene mapping of the lattice which appear to indicate groups of similar genotypes but these needed to be captured and examined in detail. The modelling process needed to include a numerical method of measuring the inter-fertility of genotypes so that potential species boundaries could be recognised, and the separate colonies marked.

Having concluded in the Literature Review 2.1.2.1 that “Sterile interbreeding therefore appears to be the best criterion available for judging speciation in sexually reproductive organisms,” mutual infertility could be considered as a good available indicator of a geographical boundary between species. It was therefore decided to devise an algorithm for marking the geographical boundaries in the lattice between potential species by tracing lines of low inter-fertility.

4.5.1.1 Hamming distance and speciation

Waga et al. [2007] claim to have identified speciation in their use of the Penna model. They used the Hamming distances\(^1\) between the genotypes of various groups as a proxy for species separation.

However, the Hamming distance does not appear to be a useful measure for identifying species separation. A small Hamming distance indicates genetic similarity but this is not, within the

\(^1\)The Hamming distance between two binary numbers is the number of differences between corresponding bits in the two numbers.
Penna model, an indicator of reproductive compatibility because similar genotypes have a significant probability of containing conflicting mutations and consequently of being incompatible.

As part of the current project, all possible pairs of 16-bit gametes were compared and the proportion of compatible pairs was evaluated for each increment of Hamming separation. This showed that at the maximum Hamming separation (16) all pairs were compatible: the pairs were all complementary and therefore had no coincident mutations. For each unit of decrease in Hamming separation the proportion of compatible pairs reduced by half. At a Hamming separation of zero the proportion of compatible pairs was \(2^{-16}\) because all the pairs are identical matches and only one pair (both zero) are compatible out of \(2^{16}\) possible pairs.

This confirms that as Hamming separation increases the proportion of compatible gametes increases so that the chance of reproductive compatibility increases. Increasing Hamming separation potentially indicates decreasing species separation. Appendix B.1.4 on page 175 gives details of this analysis.

This analysis is only a first order approximation because not all potentially conflicting mutations will be manifest in all the possible zygotes but it is a general indication that, under the rules of the Penna model, the greater the Hamming distance between genotypes the more likely they are to produce viable offspring. This means that the Hamming distance is not a good measure of incompatibility, in the Penna model, and consequently not a useful marker for speciation.

More significantly, this review demonstrated the inadequacy of the Penna model for modelling multiple species, an idea which is developed below in Section 4.6.

### 4.5.2 Measuring compatibility between plants

A key measure of inter-fertility between plants was devised, called the ‘local compatibility’ of a plant. It is assessed in relation to the surrounding plants within pollination range. It is defined as the proportion of possible cross-pollinations which would result in viable offspring.

In the case of 64-gene diploid plants, assuming one crossover in the creation of every gamete, there are 128 possible gametes from each plant. These are tested against the 128 possible gametes from the surrounding plants and the potentially viable pairings are counted. The proportion of these potential offspring which are viable is defined as the local compatibility of the central plant. The method employed to execute this calculation efficiently is described in Appendix A.3.5.2.

The 128 possible gametes which each plant can produce were divided into those which result from a crossover and those which do not. This allows the probability of crossover \(P_R\) to be included in the calculation, as follows.

Two gametes can be produced without crossover and these are the parental haplotypes \(H_0\) and \(H_1\). The probability of producing each of these is therefore \((1 - P_R)/2\). The other 126
possible gametes resulting from crossovers in 63 places between the 64 genes is $P_R/126$. These probabilities were accumulated in the local compatibility algorithm to indicate the probability of the success of each possible pair of parents.

To simplify this calculation it is assumed that not more than one crossover will occur and that no random mutations are added during the production of the gametes. These assumptions mean that the gametes considered are within the set of possible gametes but do not include all possible gametes which might be produced in more complex cases involving multiple crossovers and random mutations. Nevertheless, as a sample of possible outcomes this method will give a reasonable representation of the potential compatibility between neighbours. If more crossovers were allowed there would be some random increase and decrease in the viability of the offspring probably resulting in no substantial change in the overall local compatibility value. If random mutations were to be included these would increase the probability of the non-viability of the offspring which would result in a generally lower value local compatibility. Therefore, the algorithm used can be expected to give correct or, in more complex cases, a slightly high result.

When mechanisms for 'phenotype matching', 'assortative mating' and 'phenotype fatality' were introduced into the model, see Section 4.7, the algorithm for local compatibility was modified to allow for their effects, see Sections 4.7.1.2, 4.7.2.2 and 4.8.1.1.

### 4.5.2.1 Compatibility boundaries

Having established the local compatibility of each plant in the lattice, it was observed that in certain circumstances, lines of low local compatibility appeared which might be incipient boundaries between different species. These have been termed 'compatibility boundaries'.

The values of local compatibility in these linear zones vary considerably and so, in order to enhance the contrast between these potential boundaries and other areas, the range of values for local compatibility is normalised to extend from $C'_L_{0}$ (a low value) to 1.0. The normalised local compatibility, $C'_L$, is derived from the local compatibility, $C_L$, by the function:

\[
C'_L = \begin{cases} 
C_L - C_{L_{min}} + (1 - C'_L_{0}) & \text{when } C_L \geq C_{L_{min}} \\
0 & \text{when } C_L = 0
\end{cases}
\]

where $C_{L_{max}}$ and $C_{L_{min}}$ are, respectively, the maximum and minimum values of $C_L$ for the whole lattice and $C'_L_{0}$ is a specified lower limit for $C'_L$, which can be useful for differentiating between low, non-zero values of $C_L$ and zero values.

The value $C_b$, below which $C'_L$ is deemed to represent a compatibility boundary, was chosen arbitrarily but values in the range 0.375 to 0.675 were found to give useful results. This range of values for $C_b$ may be considered in comparison with the simplistic arrangements of 100%
compatible ($C_L = 1.0$) and 100% incompatible ($C_L = 0.0$) neighbours shown in Figure 4.2. Were a plant to be 100% incompatible with 3 of its 8 neighbours and 100% compatible with the other 5, its local compatibility would be $5/8 = 0.675$. Similarly, with 4 or 5 incompatible neighbours the values would be 0.500 and 0.375 respectively. These are simplistic scenarios but serve to illustrate the likely range of local compatibility at species boundaries before normalisation.

Figure 4.2: Simplistic examples of $C_L$ near compatibility boundaries

The selected value of $C_b$ was used across the lattice to identify locations where plants exhibited normalised local compatibility $C'_L$ below $C_b$. These lines of plants were deemed to mark the compatibility boundaries.

This measure of compatibility differs importantly from Hamming Distance because it uses the specific mechanisms of the reproductive model to measure compatibility rather than relying on the assumed relevance of the Hamming Distance which is a measure of bit-matching. As a particular example, one may consider a gamete of 64 zero bits ($Z$) and one of its binary complement ($\bar{Z}$). The Hamming Distances and Compatibilities between these are summarised in Table 4.1 on page 63 which shows, in this simple case, the profound difference between Hamming Distance and compatibility.

<table>
<thead>
<tr>
<th>Gamete 1</th>
<th>Gamete 2</th>
<th>Hamming Distance</th>
<th>Computed Compatibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Z$</td>
<td>$Z$</td>
<td>0</td>
<td>true</td>
</tr>
<tr>
<td>$Z$</td>
<td>$\bar{Z}$</td>
<td>64</td>
<td>true</td>
</tr>
<tr>
<td>$\bar{Z}$</td>
<td>$Z$</td>
<td>64</td>
<td>true</td>
</tr>
<tr>
<td>$Z$</td>
<td>$\bar{Z}$</td>
<td>64</td>
<td>false</td>
</tr>
</tbody>
</table>

Table 4.1: Hamming Distance compared with genetic compatibility in extreme, simplistic cases

The introduction of crossovers in the generation of gametes complicates the situation which was analysed and numerically evaluated as described in Appendix B.1.3 on page 172. The assessment shown in Appendix B.1.4 on page 175 specifically rejects the use (by Waga et al. [2007]) of Hamming distance as a proxy for speciation.
4.5.3 Identifying colonies

For the purposes of this project the term ‘colony’ is used to describe a contiguous group of plants enclosed within an established compatibility boundary. The compatibility boundaries in the lattice, derived as described above, do not always form closed loops and are not even necessarily continuous, but where a continuous closed boundary exists it is assumed to enclose a ‘colony’ which may be a separate species.

4.5.3.1 Numbering of colonies

An algorithm was developed which assigns a unique ‘colony identity number’ to all plants in groups which are not separated by a compatibility boundary. This process is repeated with consecutive colony numbers until all plants in the lattice have been assigned a colony number. The cells which form the boundaries themselves are assigned the colony number zero.

It is interesting to consider how such an algorithm would respond to the situation of a ring-species on a cylindrical lattice, where each sub-species is compatible with its immediate neighbour but not to more remote genotypes. It would seem likely that such a situation would result in the detection of only one ‘colony boundary’ and this would be where the ‘ring’ is closed and interbreeding is not possible. This situation has not been investigated in a simulation.

A further algorithm was devised in order to maintain the same colony identity number for a given colony as time progresses. This was based on the assumption that the geometric centroid of a colony would remain part of the same colony, over the short term, even if the colony boundary moved. In many cases these algorithms succeeded in maintaining stable colony numbering over long time periods.

Colony numbers allowed the genetic characteristics of each colony to be analysed despite the inherent genetic variation within the colony.

Figure 4.3 shows, for the same time-step, a coloured gene map (top) representing the various genotypes of the plants with different coloured pixels; a normalised compatibility map (middle) where darker greys indicate lower compatibility at potential compatibility boundaries; and a map of identified colonies (bottom) where the coloured blocks are colonies and the black lines are the compatibility boundaries. The black spots within the colonies indicate the locations of dead or immature plants for which compatibility has no meaning and which are assigned a local compatibility of zero as they are unavailable as pollination partners. This example is shown because it demonstrates a clear separation into, in this case, five colonies. Other results are not always as clearly defined as this example.

It should be noted that some genetic variations, shown by colour differences on the gene map (top), are nevertheless enclosed within the boundaries of the colonies. The validity of grouping
these plants together as a species could then be tested by collating and comparing the genetic characteristics of the colonies.

4.5.3.2 Observation of colonies

The concept of colonies was devised in the context of the original Penna model where the fertility between plants is based on testing for double mutations in potential offspring. As the model was developed further, the local compatibility algorithm was amended to allow for other selection criteria as they were introduced.

4.6 Modelling multiple species

A fundamental requirement of this research into speciation is the simulation of co-existing multiple species. Using the Penna model in its original form, it became apparent that this model might not be well-suited to modelling multiple species.

4.6.1 Penna model of one species

One criticism of the Penna model is its tendency to suppress crosses between similarly mutated parents. Where there is a mutation at the same locus in both parents there is a probability of
50\% (depending on any crossover) that the offspring will be non-viable because the mutations coincide in the zygote. This seems a major limitation to the reality of the model because, in nature, organisms with similar genomes are more likely to mate successfully than those with dissimilar genomes.

Waga et al. [2007] refer to ‘sympatric speciation’ in the Penna model but acknowledge that: “This promotes the strategy of ‘looking for complementary haplotypes’ (like bit-strings 10001 and 01110). In complementary pairs, haplotypes ‘fit’ each other and produce heterozygous loci.” There can only be successful mating within these groups where there is not crossover of chromosomes during meiosis and Waga et al. acknowledge this form of ‘speciation’ is only possible when the probability of recombination (crossover) is set low, typically 10\%. These limitations do not seem to realistically represent natural reproduction within species where there is variation of forms (varieties) within an overall consistency of phenotypic form.

Simulations of the development of genetic diversity, reported later in Section 5.1, show that in its basic form the Penna model tends to favour only one particular genotype, essentially representing one species. Further simulations which tried to establish stable multiple species (see Section 5.3) confirmed that the original Penna model is not good at representing multiple species. The age-related mutations mechanism originally conceived by Penna represents diseases of one species. The original form of the model does not seem suitable for modelling the multiple species required to investigate speciation. These ideas are demonstrated in the simulations recorded in Section 5.3.

The fundamental reason why this model is not able to simulate speciation is that it is essentially a single species model. The model is based on one optimal genotype, namely the ‘zero-genotype’ with no mutations, and all other genotypes tend to revert to this form over time, as illustrated in the diversity simulations recorded in Section 5.3. This is the way in which it has been used in the past, for example in the investigation of the onset of age-related disease by Stauffer [2007]. The zero-genotype will always produce unmutilated gametes which are compatible with all other gametes with any number of mutations because there can be no fatal double mutations in the zygote. This means that, in a mixed population, the zero-genotype will always find a mate and will therefore be evolutionarily favoured. This inherent preference for the zero-genotype is the origin of the ‘purification’ process observed by Waga et al. [2007]. There is no provision in the model for any other genotype which is as universally inter-fertile as the zero-genotype. In terms of a fitness landscape, the original model contains only one fitness peak, at the zero-genotype, and while other forms may exist they will always be transient because no other fitness peaks exist in the landscape.

If speciation is to be modelled and various species are to coexist, several genotypes must be able to persist with equal possibility of stability. In its original form the Penna model does not provide this necessary ‘level playing field’ for speciation simulations and therefore some changes were required.
Cebrat et al. [2012] on the evolution of gene grouping, claim that the groups of different genotypes which develop in their model are examples of speciation, but this interpretation is subject to doubt. The groups concerned are most likely to breed successfully when the probability of recombination is low so that gametes are formed containing only complete parental chromosomes and do not include other sequences resulting from crossovers. However, this combining of complementary gametes does not seem to adequately represent the diversity of speciation because it relies on a fertility barrier based only on the absence of crossovers.

The age-related genetic testing in the Penna model can be switched off but this would leave the model with no mechanisms for selecting one genotype over another. It therefore was necessary to add some other selection mechanisms to the model.

### 4.7 Selection mechanisms

As discussed in Section 2.2 of the Literature Review, natural processes of sympatric speciation appear to rely on various selection mechanisms. The selection mechanism inherent in the original Penna model, as discussed above, tends to favour a single genotype or species. The following sections describe the development of other selection mechanisms in the Sympatria model, specifically: ‘phenotype matching’, ‘assortative mating’ and ‘predation’.

#### 4.7.0.1 Expressed phenotypic attributes

All of these new selection mechanisms are driven by phenotypic attributes and therefore, in the context of the current simulation model, it is necessary to derive phenotypic attributes, in some reasonably realistic way, from the known genotype of an organism. The current model does not attempt to simulate the complexities of epigenetics and the influence of environmental conditions on gene expression but simply derives attributes numerically from the two chromosomes in each organism. For the purposes of this project, phenotypic attributes are assumed to be defined by the combination of the two genotypic chromosomes. Mutated genes (1s) are arbitrarily assumed to be dominant and the genes expressed in the phenotype can then be simply given by:

\[
F_i = H_{0i} \cup H_{1i}
\]  

(4.8)

where \(H_{0i}\) and \(H_{1i}\) are the two chromosomes of an organism \(i\).

This derivation of expressed phenotype genes is used throughout the model where phenotypic attributes are employed.
4.7.1 Phenotype matching

‘Phenotype matching’ was developed to provide a barrier to mating between dissimilar plants. A pattern of gene loci is defined as the ‘phenotype loci’, an input parameter, and mating pairs are required to have matching expressed genes at these loci, with some tolerance of difference. This was intended to represent the essential structural compatibility of the phenotypes which must match in parents thus modelling the natural requirement for mates of a similar phenotypic form. For example, plants need to have compatible forms of pollen, and overlapping fertility seasons.

This mechanism requires compatible parents to have matching phenotypic expressed genes at the specified phenotype loci. So that the matching does not have to be exact, a ‘phenotype tolerance’ is allowed which specifies the number of the phenotype loci at which the parents’ expressed genes may differ. This assessment of parental similarity is made before the mate is selected which implies that dissimilar individuals are ignored in the selection of mates. This is intended to simulate pre-zygotic differentiation between species and is considered separate from sexual attractiveness (assortative mating) which is described in Section 4.7.2. Assortative mating as assumed to be based on the relative attractiveness of mates whereas phenotype matching represents the absolute impossibility of mating between phenotypes because of issues of structure or seasonal timing of fertility.

If mutations accumulate at the phenotype loci in different groups of plants, these groups have the potential to become reproductively isolated from other groups with different expressed genes. The intention is to represent the long-term divergence of species which become progressively more isolated by the acquisition of special traits. There is a risk that the phenotype matching mechanism might not allow mutations to accumulate at the phenotype loci but this exclusivity is alleviated by the introduction of phenotype tolerance.

4.7.1.1 Phenotype matching mechanism

In the Sympatria model, a subset of gene loci, the ‘phenotype matching loci’ ($L_{pm}$), are defined as those important to the compatibility of phenotypes, and is expressed as a binary number with 1 at each relevant locus. The phenotype genes ($E_i$ for organism $i$) expressed at these loci are therefore given by:

$$E_i = F_i \cap L_{pm} = (H_{0i} \cup H_{1i}) \cap L_{pm}$$

(4.9)

Organisms $i$ and $j$ are considered compatible if:

$$\beta(E_i \oplus E_j) \leq T_{pm}$$

(4.10)
where:

\( \beta() \) is a bit-counting function;

\( T_{pm} \) is the specified phenotype tolerance;

\( \oplus \) is the exclusive OR operator.

If this test fails then \( i \) and \( j \) are not considered as potential mates, but if it succeeds then they will be included in the set of potential mates for selection either randomly or by assortative mating if this is being used.

4.7.1.2 Phenotype matching and local compatibility

In assessing local compatibility (see Section 4.5.2) phenotype matching is a precondition for acceptance of a mate and is a function of only the expressed genes in the phenotypes of the two potential parents. The phenotype matching test between potential mates, as described above, was therefore incorporated into the local compatibility algorithm as a precondition for the other tests of potential mates.

4.7.2 Assortative mating

The phenotype matching mechanism described above is intended to model the structural compatibility of similar organisms. Sexual selection of mates is modelled by a probabilistic method referred to here as ‘assortative mating’.

Assortative mating involves the selection of a mate which has certain attractive attributes even when other mates are equally available and potentially inter-fertile. In the current research this is taken to mean that the selection of a mate, from a set of available mates (which are mature, within range and structurally compatible), is influenced by the similarity of certain phenotypic attributes of the mates.

The difference between the PM and AM mechanisms is that phenotype matching requires the absolute rejection of ‘unsuitable’ mates (with some tolerance of differences) whereas assortative mating allows a probabilistic choice between possible mates. In AM, the discrimination between possible mates is regulated by the AM index (\( q_{am} \)), as defined in Section 4.7.2.1 on page 70. With \( q_{am} = 0 \), the choice of mates is entirely random, but with \( q_{am} > 0 \) the probability of selecting ‘attractive’ mates is increased. In this context ‘attractiveness’ is defined by the number of phenotypic gene matches between mates at specified assortative mating loci (\( L_{am} \)).

There is evidence from the natural world of attraction to mates with similar appearance, see Jiggins [2006] concerning butterflies choosing mates with similar wing markings.

In the case of plants, the mechanisms for the choice of a mate is a little obscure because plants seem to have limited freedom of action. However, pollination mechanisms can incorporate as degree of probabilistic choice. For example in flowering plants, characteristics such
as flower colour and size may attract particular pollinators which will then transfer pollen to other plants with similar structures. This is the case with *Mimulus* quoted by Charlesworth and Charlesworth [2003] as described above in Section 2.1.2 on page 24. An insect pollinator which is attracted by structure to a particular type of plant, will continue to visit other similar plants thereby selecting mates of similar appearance ‘on behalf of’ the plants. Even in plants which are not insect-pollinated, varying degrees of overlap in the timing of fertile seasons will increase the probability of cross-pollination between mates with similar phenotypic characteristics. Mechanisms of these types can therefore be considered as proxies for mating choices by plants.

### 4.7.2.1 Assortative mating mechanism

In this mechanism, the phenotypes of a pair of potential mates are tested for the number of identical expressed genes at a defined set of loci.

The phenotypic attributes relevant to assortative mating are a subset of the expressed genes $F_i$, derived in Equation 4.8. The relevant genes ($G_i$) are at the assortative mating loci ($L_{am}$) which is expressed as a binary number with 1 at each relevant locus. The genes relevant ($G_i$) are therefore given by:

$$G_i = F_i \cap L_{am}$$  \tag{4.11}$$
$$G_i = (H_0 \cup H_1) \cap L_{am}$$  \tag{4.12}$$

The assortative mating algorithm then compares the expressed genes $G_i$ in plant $i$ with those ($G_j$) in an available mate $j$. The number of matching genes is given by: $(N_{am} - \beta(G_i \oplus G_j))$, where $\beta()$ is a bit-counting function and $N_{am} = \beta(L_{am})$, the total number of AM loci.

The probability $P_i(j)$ that parent $i$ will select mate $j$ from $n$ possible mates could then be assumed to be proportional to the number of matching genes, and given by:

$$P_i(j) = \frac{N_{am} - \beta(G_i \oplus G_j)}{\sum_{k=1}^{n}(N_{am} - \beta(G_i \oplus G_k))}$$  \tag{4.13}$$

However, in order to allow the influence of assortative mating to be experimentally varied an arbitrary parameter, the ‘assortative mating index’ ($q_{am} \geq 0$), is introduced, and the probability $P_i(j)$ that $i$ will select mate $j$ from $n$ possible mates is modified to:

$$P_i(j) = \frac{(N_{am} - \beta(G_i \oplus G_j))^{q_{am}}}{\sum_{k=1}^{n}((N_{am} - \beta(G_i \oplus G_k))^{q_{am}})}$$  \tag{4.14}$$

Increasing the assortative mating index ($q_{am}$) has the effect of accentuating the relative attractiveness of different mates by increasing the probability of their selection. When $q_{am} = 0$, all mates have an equal probability of being selected and $P_i(j) = 1/n$ for all $j$ and assortative
mating is effectively switched off. When $q_{am} = 1$, $P_i(j)$ is proportional to the number of gene matches at the AM loci, as in Equation 4.13, above. As $q_{am}$ increases, the advantage of having gene matches at the AM loci is enhanced.

During input processing, the value of $q_{am}$ is limited to 100 in order to avoid any numerical overflow problems. Typically values between 1 and 8 have been used in the simulations, as reported in Section 5.3.

### 4.7.2.2 Assortative mating and local compatibility

The local compatibility algorithm (see Section 4.5.2) was modified to allow for assortative mating which, like phenotype matching, is a function of only the expressed genes in the phenotypes of the two potential parents which simplifies its calculation.

However, the role of assortative mating is fundamentally different from that of phenotype matching because assortative mating dictates the probability of selecting one mate from all possible mates in a set within the pollination range. But, by definition, local compatibility is an assessment of the total probability of a successful mating within that range and therefore relative preferences within the set of possible mates do not affect the overall local compatibility. Assortative mating only impacts local compatibility where a pair of mates has zero probability of selection, thereby removing the pair from the compatible set.

A pair of possible mates, $i$ and $j$, will have a selection probability ($P_i(j)$) of zero, so that from Equation (4.14):

$$P_i(j) = 0$$

$$\frac{(N_{am} - \beta(G_i \oplus G_j))^{q_{am}}}{\sum_{k=1}^{n}((N_{am} - \beta(G_i \oplus G_k))^{q_{am}})} = 0$$

When $q_{am} > 0$, which is required when using assortative mating, this can be simplified to:

$$N_{am} - \beta(G_i \oplus G_j) = 0$$

$$\rightarrow \beta(G_i \oplus G_j) = N_{am}$$

By definition: $N_{am} = \beta(L_{am})$

$$\rightarrow \beta(G_i \oplus G_j) = \beta(L_{am})$$

But, $G_i$ and $G_j$ are already limited to the loci $L_{am}$ (see 4.11):

$$\rightarrow (G_i \oplus G_j) = L_{am}$$

This simplifies the computation of the impact of assortative mating on local compatibility, because if equation (4.21) is true then $i$ and $j$ will never be selected as mates and are deemed to be incompatible, whereas if equation (4.21) is false there is a finite probability that they will
mate and, in terms of assortative mating criteria, they can be considered for further assessment of compatibility.

The local compatibility algorithm in the model was adapted to allow for assortative mating using Equation (4.21).

4.7.3 Predators

The model was modified to allow the plants to be selectively killed by predators but also to acquire mutations which protected them from predation. The predators themselves were not represented in the model but their effect (predation) was seen as an increase in the environmental death rate, calculated according to the phenotype which was implied by the genotype. In effect, the predators are considered as an unseen environmental selector, which cause the death rate to increase by a specified amount, the ‘predation rate’, $D_p$.

Predators are assumed to be attracted only to plants with a pattern of expressed genes $E_{Pa}$ at specified ‘predator attraction loci’, $L_{Pa}$. Plant acquire resistance to predators when they display certain phenotypic attributes which deters the predators. For example, toxicity, a bitter taste or thorns; these factors must be *indicative* so as to deter the predators before they kill the plant. These deterrent attributes are assumed to be associated with a pattern of expressed genes $E_{Pd}$ at specified ‘predator deterrence loci’, $L_{Pd}$. When the attraction characteristics are present, and the deterrents are not, then a plant has an increased probability of being killed by the predator. The predation rate, $D_{pi}$, experienced by plant $i$ is therefore given by the function:

\[
\begin{align*}
\text{If } ((L_{Pa} \cap E_i) = E_{Pa}) \cap (((L_{Pd} \cap E_i) \neq E_{Pd}) \cup L_{Pd} = 0) \\
\text{then } D_{pi} &= D_p \\
\text{else } D_{pi} &= 0
\end{align*}
\]

where $E_i$ is the gene pattern expressed in the phenotype of plant $i$, see 4.7.0.1, above.

The resulting enhanced probability of death remains constant for the life of the plant and therefore needs to be only calculated once, at birth.

The predation rate is added to the death rates: $D_0$, the general rate, and $D_e(V)$ which is a function of the ‘environmental value’ $V$ (when this is in use).

4.7.3.1 Predator-prey interaction

It should be noted that the interaction of the populations of predator and prey is not fully represented in the present model. This model only represents the population of the prey,
the plants, while the predators are ‘unseen’ actors represented only as a constant, enhanced
death-rate for prey with a particular phenotype.

The interactions of prey and predator can be analysed more fully using the Lotka-Volterra
equation, as described by Berryman [1992]. Solutions to this equation typically show a cyclical
rise and fall of the predator and prey populations out of phase with each other, and typically
neither prey nor predator is totally eliminated because one is an essential food source for
the other. However, were the prey to be eliminated then the predators would also die out if
there were no other resources on which to prey. In the current study, the ultimate fate of the
predator is immaterial provided that predation is not required to reassert itself to control a
second population of prey. It is therefore not unreasonable to assume that the prey can be
totally eliminated provided that this only happens once.

4.8 The Dobzhansky-Müller mechanism

The Dobzhansky-Müller mechanism, as described in Section 2.3.2, is a theoretical mechanism
which suggests how a sympatric population could evolve into mutually incompatible forms.
There is no evidence of this mechanism having been simulated and it was therefore interesting
to investigate this mechanism in the current model. This was done with minimal alteration to
the model to allow a ‘fatal phenotype’ to be excluded from the population, see below.

The input scenarios were extended to provide a simplify format for defining the loci $a$ & $b$, but
this was not a functional change to the simulation.

Details of simulations of Dobzhansky-Müller are given in Section 5.4.

4.8.1 Phenotype fatality

In order to simulate Dobzhansky-Müller it was necessary to introduce a criterion which aborts
offspring with a fatal combination of genetic characteristics; in this case the $AB$ combination.
This was achieved by introducing ‘phenotype fatality’ specified as a set of loci ($L_{pf}$) at which
a specified pattern of expressed genes ($G_{pf}$) in the phenotype would be fatal. This test was
applied by deducing the expressed phenotypic genes from the zygotic genes, so that the zygote
is viable if:

$$(L_{pf} \cap (H_{z0} \cup H_{z1})) \neq G_{pf}$$  (4.25)

where $H_{z0}$ and $H_{z1}$ are the two gametes forming the zygote.

In using of the model to simulate the Dobzhansky-Müller mechanism, the testing for double
mutations (Penna’s ageing process) was suppressed by not specifying any ‘critical’ loci for the
ageing process. This causes the model to omit all the ageing test procedures.
4.8.1.1 Phenotype fatality and local compatibility

In order to include the phenotype fatality in the algorithm for local compatibility it is necessary to assess the projected phenotypic attributes from all the zygotes which might arise between each pair of potential mates. This is similar to the original assessment of double mutations described in Section 4.5.2.

The local compatibility algorithm was therefore modified so that the compatibility of a pair of possible gametes is set to zero when the phenotype projected from their potential zygote would be fatal.

4.9 The complete *Sympatria* model

This chapter has summarised the stages of development of the *Sympatria* model, broadly in the order in which the developments were implemented. As far as possible the model has been made ‘backwardly compatible’ so that earlier simulation inputs would continue to work with later versions of the model. The model was continually modified throughout the research project and it was completely rewritten once.

Details of input and output mechanisms and some interesting technicalities of the code are given in the Appendices:

- A.1: Input to the model
- A.2: Output mechanisms
- A.3: Technicalities of the model

Appendix B includes some mathematical analysis of the mechanisms of the model.

The next chapter describes the simulations undertaken with the model.
Chapter 5

Simulations using the *Sympatria* model

This chapter describes simulations undertaken using the *Sympatria* model as it was developed over the course of the research project. The simulations have been divided into the following groups:

- 5.1: Simulations of the growth in genotype diversity
- 5.2: Simulations of environmentally acquired mutations
- 5.3: Simulations maintaining multiple genotypes
- 5.4: Simulations with the Dobzhansky-Müller mechanism
- 5.5: Simulations of Dobzhansky-Müller with predators

The overall simulation strategy which was developed during the project is summarised below. Initially the model was used to reproduce some of the observations of others, notably Waga et al. [2007]. In particular, simulations of the development and decline of genetic diversity were reproduced and are described in Section 5.1 on page 78.

Subsequently, a spatially varying environment was introduced into the model (see Section 5.2 on page 82) but failed to exhibit any discontinuity which might initiate irreversible genetic isolation. This approach was also thought not to model sympatric speciation because the use of a varying environment could be considered as not truly *sympatric*.

Some effort was then put into devising algorithms for defining and marking boundaries between potentially genetically incompatible groups of plants in the modelling lattice. The objective was to identify interfertile groups which might be, or might become, separate species. Identifying and demarcating these groups was difficult because of the great diversity of genotype which can arise in a simulation. These techniques were discussed in Section 4.5 on page 60 but were never systematically applied as a simulation technique. While this approach was technically
successful at showing where speciation might be occurring, it was difficult to achieve convincing speciation in the very complex populations which arise in the model.

Ultimately a different approach was used: instead of looking for speciation within a complex population, the strategy moved to modelling the Dobzhansky-Müller mechanism in which the genotypes of the different species and their compatibilities were predefined and consequently they could be identified by inspection. The investigation then concentrated on the transition between these species.

However, the simulation of sympatric speciation required a modelling environment in which multiple species could coexist which was shown not to be the case for the Penna model. As described in Section 4.6 on page 65, the Penna model, in its original form, was found to be incompatible with the need to model multiple species. This deficiency was demonstrated in the simulations described in Section 5.3.1 on page 97. The simulation strategy therefore moved away from the strict exclusion of double mutations (the Penna ageing process) and towards mechanisms which allow different mutation patterns to coexist stably in the model.

In order to maintain evolutionary selection in the model, some alternative mechanisms were required and phenotype matching and assortative mating were introduced to provide selection mechanisms when allowing multiple genotypes to coexist on equal terms. The details of these techniques are described in Section 4.7 on page 67 and simulations testing their performance are recorded in Section 5.3 on page 96.

With potential stability of multiple species built into the model, it was then possible to undertake some simulations based on the Dobzhansky-Müller (DM) mechanism, the principle of which is described in the Literature Review, Section 2.3.2 on page 31. The simulations described in Section 5.4 on page 121 tested the effect on the Dobzhansky-Müller mechanism of varying degrees of phenotype matching and assortative mating and different starting populations.

The possible influence of mimicry and predation on speciation initiated the final aspect of the research strategy, where the effect of predation on the Dobzhansky-Müller speciation mechanism was tested in two ways. Firstly, a predator was used to exclude the intermediate \((AB)\) phenotype which in the Dobzhansky-Müller mechanism is required to be non-viable. Secondly, a predator was introduced to deplete the original population thereby allowing new genotypes to prosper. The possibilities were investigated in the simulations described in Section 5.5 on page 135.

### 5.0.1 Records of detailed data

The principal input and output data for all the simulations described here are summarised in the Appendix C on page 178. The objectives and conclusions from each set of simulations are recorded in the following sections of this chapter.
In reality, numerous other simulations were run, many of which were not very interesting but these runs were useful in indicating the sensitivity of the model to various parameters. The results of many simulations are not recorded in this report but gaps in the numbering of the simulations described here indicate that there were many other runs.
5.1 Simulations of the growth in genotype diversity

The simulations described in this section examine the degree to which genetic diversity develops and is sustained while using the modelling mechanisms of the original Penna Model as described by Waga et al. [2007] and others. This development of diversity was investigated twice, firstly as a preliminary review of the behaviour of the model described below in Section 5.1.1; and secondly, as a direct comparison with the diversity which arose when trying to maintain separate species as reported later in Section 5.3 on page 96. Waga et al. [2007] demonstrated that, with a sufficient probability of recombination (≥ 40%) and given a long enough period, the plants in the model will become clear of all mutations. This is because mutations tend to inhibit the production of viable seed when crossing with other plants and therefore less-mutated parents can mate with a higher probability of success. These authors refer to this phenomenon as “purification” of the chromosome and it is a primary selection pressure inherent in this form of the model. Typically mutations are progressively removed from the ends of the chromosome to give greater and greater probability of compatibility with others.

The objective of these early simulations was principally to demonstrate that the newly developed model produced results which are consistent with the results reported by others.

5.1.1 Diversity in the Penna model

The data reported here is similar to that presented by the author at the Conference on Theoretical & Mathematical Biology in June 2011 in Cracow. It concerns the different rates of growth and subsequent decline in diversity, measured as the total number of genotypes present at any time.

5.1.2 Diversity in the Penna model: input

The number of different genotypes present was recorded for values of recombination probability ($P_R$) increasing from 0.10 to 1.00 with the mutation rate ($P_m$) set to zero in all cases so that genetic changes resulted only from recombinations. The initial population was *Uniform Random Comp*¹, which provided a single, universal genotype with random but complementary chromosomes. Consequently, initially there is one identical genotype at all locations, with a mutation at every locus on one or the other chromosome. Each configuration was run 20 times with a different random initial genotype each time.

The essential input assumptions are summarised in Table 5.1 on page 79 and fuller details of the input and output data are given in Appendix C.1 on page 178.

¹ See Appendix A.1.1.1 on page 158 for definitions of the possible initial populations.
Table 5.1: Key Inputs for growth and decline of number of genotypes

<table>
<thead>
<tr>
<th>Simulation</th>
<th>5.1.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td>Uniform random complementary</td>
</tr>
<tr>
<td>$P_R$</td>
<td>0.10, 0.20, 0.40, 0.80, 1.0</td>
</tr>
<tr>
<td>$P_m$</td>
<td>0.0</td>
</tr>
<tr>
<td>$t_{max}$</td>
<td>60000</td>
</tr>
<tr>
<td>Reruns</td>
<td>20</td>
</tr>
</tbody>
</table>

5.1.3 Diversity in the Penna model: results

An output chart was used which shows ‘diversity’ as the number of different genotypes present over time. This generally showed an initial rise in diversity followed by an asymptotic decline towards one genotype. In Section 4.6 on page 65 this effect was referred to as demonstrating that the basic model is not adequate for modelling speciation because it tends to favour the single unmutated genotype.

The results plotted in Figure 5.1 on page 79 show that for values of $P_R$ below 0.40 the number of genotypes continues to rise very slowly even after 60000 time-steps. With $P_R$ of 0.40 and above the number of genotypes initially grows but then declines. When $P_R$ is 0.8 and 1.0 the number of genotypes rises to several hundred but then decays to one, which is a universal coverage of plants with no mutations. This mechanism was recorded by Waga et al. [2007] and the ultimate loss of all mutations was described by them as "purification" of the genotype.

![Figure 5.1: Growth and decline of diversity for different probabilities of recombination ($P_R$)](image-url)
The “purification” process is faster when the recombination rate is higher, hence the more rapid effect for $P_R = 1.0$. Waga et al. concluded that the purification is dominant when $P_R$ is above 0.4 but no theoretical justification for this figure has been found. Figure 5.2 on page 80 shows the decline in the number of mutations during this “purification” process, using $P_R = 1.0$. Initially all plants have complementary haplotypes and therefore the proportion of mutations at all loci is 1.0, shown by the green line at the top. Subsequently, the proportion of mutations reduces, with the loci at the ends of the haplotypes being cleared first, shown in Figure 5.2 on page 80. Eventually there are no mutations at any locus.

Initially all plants have complementary haplotypes and therefore the proportion of mutations at all loci is 1.0, shown by the green line at the top. Subsequently, the proportion of mutations reduces, with the loci at the ends of the haplotypes being cleared first, shown in Figure 5.2 on page 80. Eventually there are no mutations at any locus.

![Figure 5.2: Purification of mutations from the ends of chromosomes as time progresses](image)

**Figure 5.2: Purification of mutations from the ends of chromosomes as time progresses**

### 5.1.4 Diversity in the Penna model: conclusions

This early part of the project served to demonstrate that the newly developed version of the model produced results similar to those from Waga et al. [2007]. The number of genotypes increased rapidly and then decayed towards one; the ultimate type being the unmutated form. The rate at which this development proceeded depended on the probability of crossover: with a higher probability of crossover resulting in a faster rise and fall in diversity.

It is not clear whether all mutations would eventually be eliminated even when the probability of crossover ($P_R$) is low. The simulation was not run for long enough to establish this but Waga et al. [2007] suggest that “purification” does not occur with values of $P_R$ below 0.4.

The progressive loss of mutations from the ends of the chromosomes, shown in Figure 5.2 on page 80, indicates that the “purification” process is driven by the crossover mechanism rather
than by general genetic drift which would be expected to result in the loss of mutations equally to all parts of the chromosome.

At this stage of the research project it was not realised that the “purification” phenomenon (to a single unmutated genotype) indicates that the original Penna model is too biased in favour of that single type to allow speciation to be modelled effectively. This inadequacy of the model is addressed more fully in Section 5.3 on page 96.
5.2 Simulations of environmentally acquired mutations

This section summarises some simulations introducing the environmental value (V), representing for example temperature, humidity or nutrition which can be varied across the lattice to simulate differing environmental conditions. The model was set up to allow phenotypic characteristics in the plants which protect them from adverse environmental conditions, as described in Section 4.4 on page 58. These characteristics were derived from the genotype.

The objective was to see if any interfertility barrier developed when the environmental or genetic parameters were varied. This investigation was divided into two parts.

Firstly, as reported in Section 5.2.2, a simple linear variation of the environmental value was applied using first a benign and then a more aggressive environment. This allowed the basic mechanics of the acquisition of protective mutations to be investigated. Secondly, reported in Section 5.2.3 on page 91, a series of simulations were run with increasing environmental gradient across the lattice. This was done to test whether any sudden breakdown in interfertility arose as the gradient increased.

5.2.1 Simulation methods for the environment

The general set-up of the environmental parameters is described first.

The lattice in all these simulations was cylindrical about an axis parallel to the rows, 128 cells long and 32 around the circumference. This shape was used because the length (128), being many times the distributing distance for pollen or seeds, was considered adequate to achieve the type of isolation required for speciation. The cylindrical shape offered a boundary-free extent in the other dimension. The lattice was divided, along the axis of the cylinder, into 3 equal\(^2\) reporting zones chosen to illustrate regions of high, medium and low V; the zones are numbered (0,0), (0,1) & (0,2), in order from highest to lowest V.

The plants all had 64 genes and tolerated no double mutations. The initial population consisted of a pair of parent plants with all chromosomes completely free of mutation, called here a Zero Pair. This arrangement allowed a maximum rate of change of the genetics because there is ample empty space for the distribution of new seeds with newly evolving characteristics. A more heavily populated initial configuration would probably have taken more time to evolve to suit the environment.

The environmental value (V) varied linearly along the axis of the cylindrical lattice from 50 to 0 making the rate of change (\(\Delta V\)) equal to \(-50/128 = -0.390625\) units per column. The minimum death rate (\(D_0\)) in every time-step for all plants was 0.01, and the death rate increased by \(\Delta D_V\) per unit of V when V was outside the ‘tolerated’ range \(V_L\) to \(V_H\). \(\Delta D_V\) was variously

---

\(^2\)Because 128 is not divisible by 3 the zones were actually 43, 43, & 42 columns wide respectively.
The presence of mutations at certain loci allowed a plant to tolerate a greater range of $V$ before the death rate began to increase. The tolerance range for $V$ increased by $\Delta V_m$ for each mutation at the specified ‘beneficial’ loci. Mutations at one group of loci ($L_H$) allowed the plants to tolerate higher $V$, and mutations at other loci ($L_L$) allowed tolerance of lower $V$. In the simulations described here, these beneficial loci were widely separated on the chromosome so that, during reproduction, they could be gained or lost independently of each other. The range of $V$ which could be tolerated with no increased death-rate was between $V_0 + \Delta V_m M_H$ and $V_0 - \Delta V_m M_L$, where $V_0$ was the ‘optimal’ environment for the plants (25 in this case) and $M_H$ and $M_L$ were the numbers of mutations present at the beneficial loci $L_H$ and $L_L$, respectively.

$L_H$ and $L_L$ both included 4 loci so that plants which developed these 4 beneficial mutations, for both higher and lower tolerance, could tolerate $V$ between 17 and 33 without suffering any increased death rate. It should be noted that because the beneficial loci $L_H$ and $L_L$ were different parts of the chromosome loci, a single plant could evolve to achieve tolerance to both directions albeit at the reproductive ‘cost’ of carrying multiple mutations which would tend to reduce inter-fertility and consequently reproductive chances because any additional mutations in the parent plant increased the probability of mutant genes coinciding in the seed zygote, making it non-viable. Because the overall range of $V$ present was 0 to 50, the range of $V$ experienced in the central zone (0,1), being the middle third of the lattice, is approximately 17 to 33, as shown below in Figure 5.3 on page 84. Consequently, with the parameters used, plants in the central zone could potentially evolve to tolerate the whole of the range of $V$ encountered there.

### 5.2.2 Simulations in a linear environment (EVLN)

Working with the parameters described above, the first simulations tested whether protective characteristics would be acquired in response to the environmental conditions.

The detailed input for these simulations is shown in Appendix C.2.1 on page 181.

#### 5.2.2.1 Simulations in a linear environment (EVLN): input

Two cases were tested, one with a benign and one with a very aggressive environment. In both cases, the rate of random mutation ($P_m$) was set to 1.0 mutation in each gamete and the probability of crossover ($P_c$) of 0.40 ensured that mutations would be created and exchanged sufficiently quickly to demonstrate the acquisition of resistance to the prevailing environment.
The general environmental death rate ($D_0$) was 0.01 and the increase in this, per unit $V$, $\Delta D_V$ was varied from: 0.005, 0.02, 0.05, 0.10, 0.20 but only the two extreme cases are discussed in detail here to illustrate the principles involved. The loci (56 to 59) in $L_H$ providing tolerance of high $V$ were at the high end of the chromosome, and the loci (4 to 7) in $L_L$ which provided tolerance of low $V$ were at opposite ends of the chromosome. The model was run for 2000 time-steps.

The key input parameters for these cases are summarised in Table 5.2 on page 85.

In each case the mutation distribution was plotted at 20 time intervals during each run. The mutation graphs in the results show the average distribution of single mutations along the 64 loci of the chromosome in three zones of the lattice. The overall population is also plotted against time. All the results shown are the average results of 20 model runs.

5.2.2.2 Simulations in a linear environment (EVLN): results

The results for the more benign environment (case 5.2.2.2.1) and the more aggressive (case 5.2.2.2.2) are discussed separately, below.

5.2.2.2.1 Results for a benign linear environment $\Delta D_V = 0.005$

With a low value of $\Delta D_V$ the environment can be considered fairly benign, having a slow
<table>
<thead>
<tr>
<th>Cases:</th>
<th>5.2.2.2.1</th>
<th>5.2.2.2.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Envir.</td>
<td>'Benign'</td>
<td>'Aggressive'</td>
</tr>
<tr>
<td>$\Delta D_V$</td>
<td>0.005</td>
<td>0.20</td>
</tr>
<tr>
<td>$V_0$</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>$\Delta V_c$</td>
<td>-0.3906</td>
<td>-0.3906</td>
</tr>
<tr>
<td>$D_0$</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>$P_R$</td>
<td>0.40</td>
<td>0.40</td>
</tr>
<tr>
<td>$P_m$</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>$L_H$ (as hex)</td>
<td>00000000000000F0</td>
<td>00000000000000F0</td>
</tr>
<tr>
<td>$L_L$ (as hex)</td>
<td>0F00000000000000</td>
<td>0F00000000000000</td>
</tr>
<tr>
<td>$t_{\text{max}}$</td>
<td>2000</td>
<td>2000</td>
</tr>
</tbody>
</table>

Table 5.2: Key input parameters for hostile environments

Influence on the development of mutations. At the extreme ends where $V = 50$ or 0, the death rate ($D$) can become 0.135 if there are no beneficial mutations, but only 0.095 if all the beneficial mutations are present.

Figure 5.4: Mutations acquired in zone $(0,0)$ of high $V$ in a benign environment ($\Delta D_V = 0.005$).

In zone $(0,0)$, $V$ was between 50 and 33 which exceeded the optimum value of 25 by at least 8. Mutations therefore developed in loci 56 to 59 to mitigate the effect of high $V$ as shown in Figure 5.4 on page 85. Even with all 4 beneficial mutations the plants could tolerate $V$ only up to 33 with no increase in death rate, so in this zone some increased death rate occurred. There were no extra mutations at the other end of the chromosome because there was no selection for...
Figure 5.5: Mutations acquired in zone (0,1) of medium V in a benign environment ($\Delta D_V = 0.005$).

Figure 5.6: Mutations acquired in zone (0,2) of low V in a benign environment ($\Delta D_V = 0.005$).
tolerance of low $V$. In zone $(0,1)$, seen in Figure 5.5 on page 86, $V$ varied from 33 to 17 which meant that with 4 mutations in both the high and low end of the chromosome these plants evolved to avoid any increase in death rate. There was clearly a selection pressure to develop mutations for toleration at both ends of the chromosome. Zone $(0,2)$ had $V$ from 17 to 0 and Figure 5.6 on page 86 shows the complementary reaction to that in zone $(0,0)$. Here all the beneficial mutations were concentrated in the loci which compensated for low $V$. Comparing the outer zones with the central, it can be seen that there was over 40% mutation in the outer zones where mitigation of the extreme environment is essential, compared with under 40% in the middle zone where the environment is less hostile. It is also noted that the level of mutations remains at about 20% at all loci where no beneficial mutations were available and no special selection pressures existed.

Figure 5.7: Rapid growth of the population to fill the lattice in a benign environment ($\Delta D_V = 0.005$).

The complete lattice was 32x128 having 4096 possible plant positions. The population chart, Figure 5.7 on page 87, shows that plants spread quickly, within 200 time-steps, from the initial pair of parent plants, to occupy almost all the available lattice despite the hostile environment. The lattice map of genotypes, Figure 5.8 on page 88, shows that most of the locations were occupied with some sparseness (shown in white) at the ends where the environment was most hostile.

5.2.2.2 Results for an aggressive linear environment $\Delta D_V = 0.200$

The environment in this case was much more aggressive with the death rate rising steeply
Figure 5.8: Typical map of gene types in a stable population $t = 240$ in a benign environment ($\Delta D_V = 0.005$).

outside the tolerance range which encouraged the rapid development of beneficial mutations. In Zone $(0,0)$, Figure 5.9 on page 88, over 60% of plants developed the mutations which counteract the locally high $V$ compared with 40% in the previous case when these mutation were less critical. It is interesting to note that at time-step 2000 the incidence of the beneficial mutation (shown as red and orange lines) is lower than the incidence earlier at around time-step 200. The reason for this ‘over-shoot’ of mutation is not apparent.

Figure 5.9: Mutations acquired in a zone $(0,0)$ of high $V$ in a aggressive environment ($\Delta D_V = 0.200$).

Figure 5.10 on page 89, shows that plants in the middle zone $(0,1)$ survived with mutation levels below 40% as was the case when $\Delta D_V$ was 0.005. This is because survival in this region could be maintained by the relevant mutations which were able to avoid the onset of the rapid increase in death rate caused by the high $\Delta D_V$. Plants in this zone were effectively protected from the increased value of $\Delta D_V$. As with zone $(0,0)$ the plants in zone $(0,2)$ required a high incidence of the mitigating mutations to survive at low values of $V$ as shown in Figure 5.11.
Figure 5.10: Mutations acquired in zone $(0,1)$ of medium $V$ in an aggressive environment ($\Delta D_V = 0.200$).

Figure 5.11: Mutations acquired in zone $(0,2)$ of low $V$ in an aggressive environment ($\Delta D_V = 0.200$).
Figure 5.12: Population growth to fill only about $1/3$ of the lattice in an aggressive environment ($\Delta D_V = 0.200$).

Figure 5.13: Map of gene types with stable population at $t = 1800$ covering about $1/3$ of the lattice in an aggressive environment ($\Delta D_V = 0.200$).
on page 89. It should be noted that the data plotted for zones (0,0) and (0,2) indicate that there were live plants within the outer zones, but the irregularity of the plots indicates a high random error caused by a small population which is shown in Figure 5.12 on page 90.

In this hostile environment the population was reduced in the outer regions, and the numerical data shows that typically the live population remained at about 1135, occupying some 28% of the possible 4096 locations, meaning that less than a third of the environment was colonised at any one time. The plants were concentrated in the central region as confirmed by the genotype map in Figure 5.13 on page 90 showing the extent of plants after 1800 time-steps. Only 50 out of 128 columns (39%) of locations had any plants in them. The central zone had 42 columns in it, meaning that only on the inner border of the outer zones (about 4 columns wide) were there any survivors.

5.2.2.3 Simulations in a linear environment (EVLN): conclusions

These results demonstrate that characteristics can be acquired to protect against a hostile environment. This process is not unexpected and is well understood as an essential part of natural selection. Given a hostile environment it would be very surprising to find that protective genetic characteristics did not arise by natural selection.

5.2.3 Simulations in a changing environment (EVVR)

Using a similar environment and lattice, the next simulations were designed to establish if there is any sudden change in compatibility between groups of plants as the environmental gradient becomes greater.

5.2.3.1 Simulations in a changing environment (EVVR): input

Full details of the input parameters used in these runs are shown in the Appendix C.2.2 on page 183. This simulation was conducted with separate runs for different values for $\Delta V_c$, each starting from a standard position. The key input parameters are summarised in Table 5.3 on page 92, below:

5.2.3.1.1 Quarter-column Compatibility

In this simulation quarter-column compatibility ($C_Q$) is used to assess the probability of reproductive compatibility between plants in typical columns from areas of high and low $V$. Appendix A.2.2.10 on page 162 gives more details of this output technique. The compatibility, $C_Q$, was recorded at regular time intervals during runs with different values of $\Delta D_V$ gradient.
### Table 5.3: Key input parameters values for uniform and varying environmental value

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Case 5.2.3.2</th>
<th>Case 5.2.3.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Death per environmental value</td>
<td>$\Delta D_V$</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Central environmental value</td>
<td>$V_0$</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Gradient of environmental value</td>
<td>$\Delta V_c$</td>
<td>0.0</td>
<td>0.0 to 1.0</td>
</tr>
<tr>
<td>General environmental death rate</td>
<td>$D_0$</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>Probability of recombination</td>
<td>$P_R$</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>Probability of additional mutation</td>
<td>$P_M$</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>Total time-steps</td>
<td>$t_{max}$</td>
<td>50000</td>
<td>30000</td>
</tr>
</tbody>
</table>

#### 5.2.3.2 Results - Long-term effect of zero gradient, $\Delta V_c$

In this preliminary simulation, the gradient $\Delta V_c$ was set to zero which has the effect of making $V$ uniform across the whole lattice and equal to the optimal value $V_0$. The model was then run for 50000 time-steps and repeated 10 times. The objective was to see whether, and how quickly, the compatibility between different parts of the lattice ($C_Q$) became stable in this base case.

![Figure 5.14: Compatibility ($C_Q$) over time for uniform environmental value ($V$)](image)

The graph in Figure 5.14 on page 92 and data in Table 5.4 on page 93 show $C_Q$ against time for $\Delta V_c = 0$. The data shows that $C_Q$ stabilises to approximately 0.60 after about 10000 time-steps indicating that, in the larger simulations, using various values of $\Delta V_c$, the populations would
<table>
<thead>
<tr>
<th>Time</th>
<th>0.0</th>
<th>2.5</th>
<th>5.0</th>
<th>7.5</th>
<th>10.0</th>
<th>12.5</th>
<th>15.0</th>
<th>17.5</th>
<th>20.0</th>
<th>22.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_Q$</td>
<td>0.76</td>
<td>0.64</td>
<td>0.61</td>
<td>0.61</td>
<td>0.60</td>
<td>0.61</td>
<td>0.60</td>
<td>0.58</td>
<td>0.60</td>
<td>0.61</td>
</tr>
<tr>
<td>25.0</td>
<td>27.5</td>
<td>30.0</td>
<td>32.5</td>
<td>35.0</td>
<td>40.0</td>
<td>42.5</td>
<td>45.0</td>
<td>47.5</td>
<td>50.0</td>
<td></td>
</tr>
<tr>
<td>0.61</td>
<td>0.60</td>
<td>0.59</td>
<td>0.60</td>
<td>0.59</td>
<td>0.60</td>
<td>0.60</td>
<td>0.59</td>
<td>0.61</td>
<td>0.60</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.4: Compatibility ($C_Q$) over time for uniform environmental value ($V$) (Time is in 1000's). Case 5.2.3.2

become stable within, say 30000 time-steps.

In selecting the maximum useful value for $\Delta V_c$ consideration is given to the width of the ‘habitable’ zone of the lattice where $D < 1.0$. As $\Delta V_c$ increases, assuming no beneficial mutations are present, consider a distance $x_H > x_0$ where $x_0$ is the middle column:

$$D(V) = D_0 + \Delta D_V (V - V_0)$$
$$V = V_0 + \Delta V_c x_H$$
$$1.0 = D_0 + \Delta D_V \Delta V_c (x_H - x_0)$$

$$x_H = x_0 + \frac{1.0 - D_0}{\Delta D_V \Delta V_c}$$

Similarly on the side of low $V$:

$$x_L = x_0 - \frac{1.0 - D_0}{\Delta D_V \Delta V_c}$$

Giving a width (in columns) of the ‘habitable’ zone as $w_c$:

$$w_c = x_H - x_L$$
$$w_c = 2 \frac{1.0 - D_0}{\Delta D_V \Delta V_c}$$
$$\Delta V_c = 2 \frac{1.0 - D_0}{\Delta D_V \cdot w_c}$$

(5.1)

The whole width of the lattice (128) is ‘habitable’ until:

$$w_c = 128$$

$$\Rightarrow \Delta V_c = 1.55$$

(5.2)

This indicated that when $\Delta V_c = 1.0$ the whole of the lattice is still theoretically ‘habitable’ but statistical variation will result in some sparse population in the end of the lattice.
5.2.3.3 Simulations in a changing environment (EVVR): results

In this case a range of $\Delta V_c$ was tested to see the effect of high values. In previous simulations $V$ typically varied from 0 to 50 across 128 columns in the lattice which is equivalent to a $\Delta V_c$ of 0.39 and gave useful results and, as shown above, a value of 1.0 would not be so extreme as to eliminate the population completely. It was therefore decided to use a range of $\Delta V_c$ between 0.0 and 1.0 to test for the changes in $C_Q$.

![Figure 5.15: Variation of compatibility ($C_Q$) with gradient ($\Delta V_c$) of the environmental value. Case 5.2.3.3](image1)

![Figure 5.16: Gene map at time 15000 with gradient $\Delta V_c = 1.0$ showing sparse population in hostile areas. Case 5.2.3.3](image2)

In Figure 5.15 on page 94, it can be seen that $C_Q$ fell steadily as the gradient rose from 0.0 to about 0.5. With $\Delta V_c$ above 0.5 the $C_Q$ rose slightly but remained substantially constant. Figure 5.16 on page 94 shows the lack of population at the extremes in the case with $\Delta V_c = 1.0$ and this sparsity of population was probably the reason for the increased irregularity of the $C_Q$ results, for high values of $\Delta V_c$, due to a high random error caused by a small surviving
population. The results also show no clear speciation although the compatibility between the two sides of the lattice is halved from about 0.6 to about 0.3 by the increase in $\Delta V_c$ from 0.0 to 1.0.

5.2.3.4 Simulations in a changing environment (EVVR): conclusion

These results demonstrate that the system responds to a varying environmental gradient ($\Delta V_c$) which is consistent with expectations. However, there is no indication of a sudden shift in compatibility at any particular value of environmental gradient.

A different approach seemed to be needed to create conditions which will allow speciation to occur.

5.2.4 Environmental variation and sympatric speciation

All the results using a varying environment have not demonstrated any tendency to speciation. There is no apparent separation into non-interfertile groups which might be considered as different species. Furthermore, the spatial variation of the environment should probably not be interpreted as truly sympatric. This is a semantic point which deserves some more discussion.

As stated in 3.2.2, for the purposes of the current project, a strict definition of ‘sympatric’ has been used: it is assumed that the same environmental conditions must apply to the whole of an area for it to be considered ‘sympatric’. The issue of what is truly sympatric remains a difficult question. It could be argued that different life-experiences of two adjacent organisms is an allopatric scenario but this argument would make a sympatric scenario almost impossible to achieve. Alternatively an environment across which conditions vary gradually, with no defined barrier, could be considered sympatric which would give ‘sympatric’ a very wide definition.

The reason for choosing a strict definition of ‘sympatric’ is to exclude scenarios where there is gradual change. It seems clear that a mountain slope extending from low-lying temperate conditions, to high-lying cold conditions is not sympatric even though there is no defined barrier on the slope. A sympatric situation would seem to require the same conditions throughout, and on this basis a varying environment (as modelled in the above simulations) cannot be considered sympatric.
5.3 Simulations maintaining multiple genotypes

An essential requirement of sympatric speciation modelling is to be able to maintain different genotypes simultaneously in the model so that their interaction and possible separation can be tested. After using the model in its original form, it seemed that the model was not able to support the co-existence of multiple species.

The simulations described in this section were therefore undertaken to investigate the proposition that the original model cannot support multiple species and to test other modelling mechanisms which might be better suited to investigating sympatric speciation. The three approaches can be summarised as follows:

- Penna’s original using *zygote testing* as in the earlier simulations in Section 5.1 on page 78;
- *Phenotype matching* (PM) which allows only similar phenotype to mate, as described in Section 4.7.1 on page 68;
- *Assortative mating* (AM) which allows sexual selection from available mates, as described in Section 4.7.2 on page 69.

The simulations described here were not expected to generate multiple species but were simply intended to indicate if multiple genotypes could survive for a substantial period of time. The persistence of genotypes was measured by counting the number of different genotypes present in the population as time progressed. This technique does not directly measure the persistence of any one genotype, and genotypes may come and go over time, but it does give a general indication of whether the system supports simultaneous multiple genotypes.

In the following cases the model was started in a standard condition (as detailed in Section C.3 on page 184). The three initial population configurations used in these cases, known as *UniformRandomComp*, *UniformRandomZero* and *RandomZero* which are defined as follows:

*UniformRandomComp*: the whole lattice is set with the same random chromosome paired with its complement.

*UniformRandomZero*: the whole lattice is set with the same random chromosome paired with a zero chromosome.

*RandomZero*: every cell in the lattice has a different random chromosome paired with a zero chromosome.

A full list of possible initial population configurations is given in Section A.1.1.1 on page 158.
5.3.1 Genotype diversity with zygote testing

The first simulations (cases: FZG0, FZG1, FZG2) investigated the development of genotype diversity in the original Penna model under zygote testing. The term zygote testing is used here to mean the basic mechanism of the Penna model which results in the death of any zygote with more than the specified tolerance of double mutations at the specified critical gene loci; death may occur when the zygote forms or later in life if ageing is applied.

5.3.1.1 Genotype diversity with zygote testing: input

In these cases, all gene loci were considered critical and no double mutations were tolerated, i.e. birth criticality \( X_b \) was set to 64, and tolerance of double mutations \( T_m \) to zero.

In case FZG0 the probability of recombination \( P_R \) was set to 0.2, 0.4 and 0.8. The simulations were run for 15000 time-steps and repeated 20 times. A subset of these scenarios were run over longer periods as cases FZG1 and FZG2 to verify the ultimate outcome over extended time periods of up to 150000 time-steps.

Full details of all the inputs are shown in Section C.3.1.1 on page 184.

5.3.1.2 Genotype diversity with zygote testing: results

The results showed that, generally, after 15000 time-steps the whole population had reverted to a single genotype with no mutations. The speed of reaching this state increased with increasing probability of recombination \( P_R \) which is expected because changes are distributed across the population faster with a higher recombination rate. This also concurs with the findings of Waga et al. [2007] and is similar to results of the growth in genotype diversity recorded in Section 5.1 on page 78 of this report.

The time taken to reach uniformity of genotype is also affected by the configuration of the initial population with the Random Zero cases losing mutations most quickly and the Uniform Random Zero cases taking the longest time. The initial Random Zero populations were essentially unsustainable because, having one random chromosome and one zero, they were unlikely to provide viable mates except by pairing zero gametes together; the population therefore was very quickly replaced by unmutated (zero-zero) genotypes. The Uniform Random Comp case cleared the mutations the most slowly and when the recombination probability was 0.2 and 0.4 mutations were not cleared until 50000 and 150000 time-steps respectively. These latter two cases (FZG1 and FZG2) therefore had to be run over extended time-scales.
5.3.1.3 Genotype diversity with zygote testing: conclusions

These outcomes confirm that given enough time the Penna model, using zygote testing, will revert to a population of unmutated individuals, described by Waga et al. [2007] as the ‘wild’ state.

These results confirm the suggestion that the original form of model is essentially a model of one species with the possibility of occasional ‘disease’ mutations. This characteristic seems to make the model, in its original form, unsuitable for the investigation of multiple species and, therefore, not suitable for investigating sympatric speciation.

However, if the zygote testing were to be abandoned the model would proliferate diverse genotypes without constraint. It was therefore necessary to incorporate other selection mechanisms into the model which, while allow diversity, would also tend to favour mating between similar phenotypes in order to enhance the development of colonies with similar genotypes.

5.3.2 Testing of phenotype matching

The following simulations were designed to test the ability of the model to retain mutated genotypes using a mechanism called here Phenotype Matching (PM). This allows mating only between parents with similar phenotype characteristics represented by a specified part of the chromosome. Details of how this was implemented in the model and how phenotype characteristics are evaluated are given in Section 4.7.1 on page 68.

The principal inputs for this mechanism are:

- Phenotype loci: which define the parts of the chromosome to be compared, and
- Phenotype tolerance: which defines the number of genetic mismatches allowed at these loci.

In the model, phenotype matching could have been used in conjunction with Penna’s zygotic testing but that complication was avoided and phenotype matching has been used alone here, with varying tolerance of phenotypic mismatching.

5.3.3 Preliminary test of phenotype matching (PT03)

These preliminary simulations were intended to test that phenotype matching (PM) was able to maintain a non-zero chromosome over a long period of time, in this case 800000 time-steps. In order to test the extent to which PM would ‘protect’ the original genotype, mutations were allowed at the rate of 1 per meiosis but PM tolerance of 8 was allowed which meant that mates could have up to 8 mismatches of phenotypically expressed genes at the specified loci.
It should be noted that the 800000 time-step run was much longer than the 20000 to 60000 periods generally used in this project.

Full details of the input and results from these simulations can be found in Section C.3.2 on page 192.

5.3.3.1 Preliminary test of phenotype matching (PT03): input

As a preliminary test an extremely varied mutation pattern was used with the mutations \texttt{0x0F0F0F0F0F0F0F0F} on one chromosome and no mutations on the other. The whole chromosome was designated as ‘phenotype loci’ ($L_{pm}$) which are critical to the matching process. A matching tolerance ($T_{pm}$) of 8 was allowed. The test was run for 80000 time-steps and repeated 10 times starting from the same conditions each time. The objective was to see if the mutation pattern would be preserved over a long period.

Fuller details of the input data can be seen in Section C.3.2.1 on page 192.

5.3.3.2 Preliminary test of phenotype matching (PT03): results

Figure 5.17 on page 100 shows the proportion of plants with a mutation against each gene locus as time progresses.

The results show the the development of change in the mutation pattern was very slow. After 40000 time-step new mutations were found in less than 10% of plants. Even after 800000 time-steps less than 40% of plants had acquired mutations and the general shape of the genome was maintained.

5.3.3.3 Preliminary test of phenotype matching (PT03): conclusions

In general it appears that this mechanism was successful at maintaining the genotype over a long period but allows some variation because some gene mismatching was allowed.

5.3.4 Phenotype matching with tolerance of difference (FPT0)

Having established that phenotype matching (PM) can sustain a mutated genotype over a long period, further simulations were undertaken using different degrees of tolerance to phenotypic differences. The objective was to ascertain what effect this parameter has on the persistence of different genotypes.

Details of the input and results from these simulations can be found in Appendix C.3.3 on page 194.
Figure 5.17: Stable gene patterns maintained over a long period using phenotype matching (FPT0)
5.3.4.1 Phenotype matching with tolerance of difference (FPT0): input

In the initial set of simulations (case FPT0) the whole chromosome was set as phenotype matching loci ($L_{pm} = 0$) and tolerance of mismatches ($T_{pm}$) was variously set at 0, 1, 2, 4, 8, meaning that between 0 and 8 gene differences would be tolerated when comparing the phenotypes of two potential parents. Other parameters were the same as for the previous simulations with a range recombination probabilities ($P_R$) being 20%, 40% and 80% and the same three initial populations. The probability ($P_m$) of additional mutations arising during meiosis was set to zero.

5.3.4.2 Phenotype matching with tolerance of difference (FPT0): results

More results are shown in Appendix C.3.3 on page 194.

Starting from the *Uniform Random Comp* population, which has identical genotypes in the whole population with one randomly mutated chromosome and the other its complement, and with probability of recombination ($P_R$) of 20%, the number of genotypes generally initially increased but subsequently reduced, as illustrated in Figure 5.18 on page 101.

![Figure 5.18: Longterm genetic diversity from Uniform Random Complement, increases with phenotype matching tolerance ($T_{pm}$) Recombination $P_R = 20\%$ (FPT0)](image)

Figure 5.18: Longterm genetic diversity from *Uniform Random Complement*, increases with phenotype matching tolerance ($T_{pm}$) Recombination $P_R = 20\%$ (FPT0)
When zero tolerance of phenotypic difference \( T_{pm} = 0 \) was applied, the population stabilised into only two (homozygous) genotypes each with pairs of identical chromosomes; this situation was reached typically after between 3000 and 9000 time-steps. The chromosomes of these stable homozygous genotypes are derived from the complementary pair in the original population. That is to say, where the original Uniform Random Comp population all carried complementary chromosomes \( AA \), the ultimate two homozygous genotypes were \( AA \) and \( \bar{A}A \).

This outcome is not unexpected because, with no additional mutations being introduced during reproduction (i.e. \( P_m = 0 \)), the only available gametes are recombinations of the original complementary pair \( (AA) \) of which \( AA \) and \( \bar{A}A \) will be the most common zygotes, although zygotes composed of crossed gametes will also arise. The homozygous genotypes will, themselves, always produce gametes of \( A \) and \( \bar{A} \) respectively (even after crossovers) which will always exactly match their parents’ and siblings’ gametes, and be compatible, even with zero tolerance of phenotype differences. These processes are illustrated by the appearance, in the initial period of each model run, of up to 20 other genotypes which were then eliminated.

When some tolerance of phenotype difference was allowed with \( T_{pm} = 1, 2, 4 & 8 \), up to 100 genotypes were still present after 15000 time-steps. Starting with a Uniform Random Zero population, in which all plants initially have the identical genotype \( R0 \), having one random chromosome \( (R) \) and one unmутated chromosome \( (0) \), the population again converged towards two homozygous genotypes: one with two copies of the original random chromosome \( (RR) \), and the other with two copies of the unmутated form \( (00) \). But unlike in the Penna model, the unmутated population \( (00) \) was not necessarily in the majority.

This is illustrated in Figure 5.19 on page 103.

In both the Uniform Random Comp and Uniform Random Zero starting populations, values of \( T_{pm} \leq 4 \) resulted in an ultimate decline in diversity but with \( T_{pm} = 8 \) greater phenotypic difference is tolerated and diversity seemed to increase without limit. The exact value of \( T_{pm} \) which initiates this change in performance has not been examined.

With an initial Random Zero population, where a different random chromosome is placed in each plant, paired with zero, the population generally died out completely in a short time period. This is expected because with totally random neighbours each individual is unlikely to find a mate with adequate phenotype matching to be compatible. In this case the genes ‘expressed’ in the phenotype will be derived only from the random chromosome of the genotype because the other (zero) chromosome has no influence on the phenotype because (within the scope of the current model) mutations are always considered as dominant.

### 5.3.4.3 Phenotype matching with tolerance of difference (FPT0): conclusions

These results show that by using phenotype matching a number of genetic types can remain stable over long periods which was not possible with the exclusion of double mutations in the
Figure 5.19: Longterm genetic diversity from *Uniform Random Zero*, becomes unstable with high phenotype matching tolerance \((T_{pm})\) Recombination \(P_R = 20\%\) (FPT0)
original Penna model.
The coexistence of \(RR\) and \(00\) genotypes found in the case starting from \(Uniform\ Random\ Zero\) genotypes, shows that there is no apparent bias against non-zero genotypes (e.g. \(RR\)) as there was in the original Penna model.

5.3.5 Phenotype matching with various mutation rates (FPT1)

Further similar simulations were undertaken, in case FPT1, using different rates of additional mutation, \(P_m\); this parameter controls the probability of mutations being introduced in the meiosis process. The objective was to ascertain what effect this parameter had on the stable maintenance of multiple genotypes.

Details of the input and results from these simulations can be found in Section C.3.4 on page 200.

5.3.5.1 Phenotype matching with various mutation rates (FPT1): input

These simulations, similar to case FPT0, were repeated with the whole chromosome designated as phenotype matching loci \(L_{pm} = 0\) and recombination probability \(P_R\) set at 20\%, 40\% and 80\%. However, in case FPT1, the probability \(P_m\) that one new mutation will arise during every meiosis, was set to 1\%, 10\%, 50\% and 100\%. In order to avoid an excessive number of parameter combinations the tolerance of phenotypic difference \(T_{pm}\) was limited to zero in all cases.

5.3.5.2 Phenotype matching with various mutation rates (FPT1): results

The diversity of genotypes was generally greater than in case FPT0, where no additional mutation had been allowed, as one might expect because new mutations are being introduced during meiosis in addition to the distribution of mutations resulting from recombination.

This is shown in Figure 5.20 on page 105.

With mutation probability \(P_m\) of 100\% the number of genotypes tended to rise continuously over time.

With a \(Uniform\ Random\ Zero\) initial population (i.e. with genotype \(R0\) throughout the lattice) and mutation probability \(P_m\) of 1\% or 10\% the number of genotypes stabilized at between 12 and 35 after about time-step 7500. The growth and decline of diversity is shown in Figure 5.21 on page 105.
Figure 5.20: Genetic diversity over time from *Uniform Random Complement* with continual mutation (FPT1)

Figure 5.21: Genetic diversity over time from *Uniform Random Zero* with continual mutation (FPT1)
5.3.5.3 Phenotype matching with various mutation rates (FPT1): conclusions

The introduction of additional mutations \((P_m > 0)\) had the expected effect of increasing the rate of increase in diversity but did not disrupt the performance of the phenotype matching mechanism.

Overall these results of FPT0 and FPT1 show that by using phenotype matching a number of genetic types can remain stable over longer periods than with the zygotic testing used in the original Penna model.

The objective of the phenotype mechanism was to model the tendency of organisms to mate successfully with those of similar phenotypic form and this objective seems to have been met because groups of similar types have remained stable in the modelled populations.
5.3.6 Simulations with only assortative mating (FAM0)

The following simulations were designed to test the effectiveness of assortative mating (AM) in enhancing the cohesion of colonies within the model. Initially, in FAM0, assortative mating was tested alone to see whether stable groups of similar genotypes would form from a randomised population. Subsequently it was tested in conjunction with phenotype matching.

The implementation of assortative mating is described in Section 4.7.2 on page 69, and it is intended to represent the probabilistic choice of mates which have already satisfied the criterion for phenotype matching. The difference between the selection mechanisms is that phenotype matching requires the absolute rejection of ‘unsuitable’ mates (with some tolerance of differences) whereas assortative mating allows a probabilistic choice between possible mates. In AM, the discrimination between possible mates is regulated by the AM index ($q_{am}$), as defined in Section 4.7.2, varying from entirely random choice with $q_{am} = 0$, and increasing the probability of selecting ‘attractive’ mates as $q_{am}$ increases. In this context ‘attractiveness’ is defined by the number of phenotypic gene matches between mates at specified assortative mating loci ($L_{am}$).

Details of the input and results from these simulations can be found in Section C.3.5 on page 206.

5.3.6.1 Simulations with only assortative mating (FAM0): input

The whole chromosome was designated as assortative mating loci ($L_{am} = \bar{0}$), and the assortative mating index ($q_{am}$) was set to 1, 4 and 8. Recombination probability ($P_{R}$) was set to 20%, 40% and 80%. Starting populations were Uniform Random Comp (URC), Uniform Random Zero (URZ) and Random Zero (RZ). Scenarios were run, as in phenotype matching, for 15000 time-steps and repeated 20 times.

5.3.6.2 Simulations with only assortative mating (FAM0): results

In all cases where $q_{am} = 1$, the diversity of the population approached 2500 genotypes (the total population size) irrespective of the recombination probability $P_{R}$, as illustrated by the red line in Figure 5.22 on page 108. This shows AM having little influence on controlling mate selection and diversity consequently increasing.

With $q_{am} = 4$ and 8 (blue & green lines in Fig 5.22), the diversity rose initially and then decayed, the maximum diversity reached being greater with increasing $P_{R}$ and with decreasing $q_{am}$. These results were the same irrespective of whether the starting population was URC or URZ. These results are to be expected because increasing $P_{R}$ increases the mixing of genes before the assortative mating begins to reduce the diversity in the population by selection of
Figure 5.22: Genetic diversity over time from *Uniform Random Zero* with various AM index. \( P_R = 20\% \) (FAM0)
similar mates. Increasing $q_{am}$ enhances the selection effect therefore limiting the diversity more quickly.

The results in Figure 5.23 on page 109 starting from Random Zero (RZ) show different effects because the population starts with random diversity in one chromosome and no mutations in the other. In this case, when $q_{am} = 1$ (red line), the diversity generally stays high because AM has little influence. This represents unselective mating. When $q_{am}$ is 4 or 8 (blue & green lines) the diversity falls. However, this plot does not show that in these cases the overall population is also falling because few, if any, suitable mates are available.

![Figure 5.23: Genetic diversity over time from Uniform Random Zero with various AM index. $P_R = 40\%$ (FAM0)](image)

The changes in diversity for all the input combinations for this simulation are shown in Figures C.16 to C.18 of Appendix C.3.5.2 on page 206.

5.3.6.3 Simulations with only assortative mating (FAM0): conclusions

These results indicate that assortative mating can reduce diversity within the population by driving evolution towards a consistent phenotype. Assortative mating is therefore likely to be a useful mechanism in the modelling of speciation.
Further investigation was therefore planned to test the combined effect of phenotype matching and assortative mating.

5.3.7 PM and AM with initial URZ (PTAM5)

These simulations used phenotype matching (PM) and assortative mating (AM) together in order to investigate whether they are likely to preserve genotype groups better than the original Penna model did.

Details of the input and results from this simulation are shown in Section C.3.6 on page 211.

5.3.7.1 PM and AM with initial URZ (PTAM5): input

The key input values used in this case are listed below in Table 5.5 on page 110.

<table>
<thead>
<tr>
<th>Run Ref.</th>
<th>PTAM5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lattice size</td>
<td>Rows x Cols</td>
</tr>
<tr>
<td>Recombination</td>
<td>$P_R$</td>
</tr>
<tr>
<td>Mutation</td>
<td>$P_m$</td>
</tr>
<tr>
<td>PM loci</td>
<td>$L_{pm}$</td>
</tr>
<tr>
<td>PM tolerance</td>
<td>$T_{pm}$</td>
</tr>
<tr>
<td>AM loci</td>
<td>$L_{am}$</td>
</tr>
<tr>
<td>AM index</td>
<td>$q_{am}$</td>
</tr>
<tr>
<td>Initial pop.</td>
<td>URZ</td>
</tr>
<tr>
<td>Loci reported</td>
<td>$L_g$</td>
</tr>
</tbody>
</table>

Table 5.5: Summary of assumptions case PTAM5

The PM and AM loci (shown above as hexadecimal numbers) are centrally placed on the chromosome (with zeros elsewhere) and are chosen to be close together with PM loci A5 (hexadecimal) and AM loci 5A (hexadecimal) represent the binary numbers 10100101 and 01011010, respectively, which are inter-leaved on the centre of the chromosome. These close-linked central loci were selected for PM and AM loci to see if the selection mechanisms were successful in working closely together. As described by Waga et al. [2007], genes in these central loci tend to be protected from recombination ‘purification’ which generally progresses from the ends of the chromosomes.

The URZ population is the Uniform Random Zero option in which all plants initially have the identical genotype $R0$, comprising one random chromosome ($R$) and one unmutated chromosome (0).
5.3.7.2 PM and AM with initial URZ (PTAM5): results

Figure 5.24 on page 111 shows the diversity starting from the single initial genotype, rising to over 80 genotypes within 1000 time-steps then decaying to less than 10 genotypes.

![Graph showing genetic diversity over time with PM and AM operating, showing rise in unmutated genotype. (PTAM5)](image)

For clarity, this report recorded only the mutation in the central 16 loci (4 hexadecimal digits). There may also be mutations at other loci but the central 16 are of particular interest for assessing the effectiveness of the PM and AM mechanisms which have been set (by $L_{pm}$ and $L_{am}$) to focus on these central loci.

When averaged over the 10 reruns the percentage of unmutated genotypes was about 37% and this average remained between 34% and 40% from time-step 2000 onwards (see Figure 5.24) which indicates that the unmutated genotype is not becoming dominant.

Figure 5.25 on page 112 shows the development of the population averaged over 10 runs: the original form $R0$ rapidly dies out, the homozygous forms $RR$ and 00 persist and various other forms make up the remainder of the population, but this proportion is diminishing.

The average populations of the $R0$, $RR$ and 00 genotypes over time are shown in Figure 5.25 on page 112, in which the part of the population marked ‘other’ includes various genotypes.
Figure 5.25: Populations of three main genotypes over time, with PM and AM operating (PTAM5)

The percentages are relative to the maximum population, 2500 (50x50), and there is a constant vacancy of about 1% due to the general 1% death rate. It can be seen that the overall population is not in decline.

Table 5.6 on page 113 shows that only four genotypes remain after 20000 time-steps for a particular but typical run. The genotype at the start of this simulation was the random chromosome (R), in this case 00000003a0000000, paired with 0. This example is taken from rerun 009, the last of the 10 reruns of this simulation. The genotype references R and 0 refer to the original random and zero chromosomes in the population, while A and B are later mutant forms.

Three of these four genotypes are homozygous: 66% of the population is RR, containing two copies of the original random chromosome (633B), 22% is AA containing two copies of 6330 which is a derivative of the original random chromosome, 12% is 00 containing two unmutated chromosomes, one individual (RB) contains E33B which is a mutant form derived from the original chromosome 633B and the original R0 genotype has completely disappeared. The table shows the outcome in the centre of the chromosome (expressed as hexadecimal).

The homozygous genotype (6330) being 22% of the population seems to have become a stable colony of a new genotype. Analysis of the PM compatibility with the other genotypes present shows some interesting results. The three genotypes (0000), (633B) and (6330), being homologous, will express all their genes in their phenotypes (see Equation 4.8 on page 67). Phenotype matching operates, in this case, on the central loci F5AF (hex) so that expressed genes at these loci for each genotype (from Equation 4.9 on page 68) are given in Table 5.7 on page 113.

---

Section 4.7.1.1 on page 68 describes the PM mechanism.
Table 5.6: Genotype populations at time 20000 (run PTAM5-009)

<table>
<thead>
<tr>
<th>Genotype ref.</th>
<th>Population</th>
<th>Population (% of 2500)</th>
<th>Centre of chromosomes (as hexadecimal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RR</td>
<td>1639</td>
<td>66%</td>
<td>..633B.. &amp; ..633B..</td>
</tr>
<tr>
<td>AA</td>
<td>549</td>
<td>22%</td>
<td>..6330.. &amp; ..6330..</td>
</tr>
<tr>
<td>00</td>
<td>294</td>
<td>12%</td>
<td>..0000.. &amp; ..0000..</td>
</tr>
<tr>
<td>RB</td>
<td>1</td>
<td>0.04%</td>
<td>..633B.. &amp; ..E33B..</td>
</tr>
<tr>
<td>R0</td>
<td>0</td>
<td>0%</td>
<td>..633B.. &amp; ..0000..</td>
</tr>
</tbody>
</table>

Table 5.7: PM expressed genes (run PTAM5-009)

<table>
<thead>
<tr>
<th>Genotype ref.</th>
<th>R0</th>
<th>00</th>
<th>RR</th>
<th>AA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chromosomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H₀</td>
<td>633B</td>
<td>0000</td>
<td>633B</td>
<td>6330</td>
</tr>
<tr>
<td>H₁</td>
<td>0000</td>
<td>0000</td>
<td>633B</td>
<td>6330</td>
</tr>
<tr>
<td>Phenotype genes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fᵢ</td>
<td>633B</td>
<td>0000</td>
<td>633B</td>
<td>6330</td>
</tr>
<tr>
<td>PM loci</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lₚₘ</td>
<td>F5AF</td>
<td>F5AF</td>
<td>F5AF</td>
<td>F5AF</td>
</tr>
<tr>
<td>PM genes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eᵢ</td>
<td>612B</td>
<td>0000</td>
<td>612B</td>
<td>6120</td>
</tr>
</tbody>
</table>

Table 5.8: PM mismatched genes (run PTAM5-009)

In this simulation the tolerance of PM differences ($Tₚₘ$) was set to zero, which means that in Table 5.8 all the genotypes with more than zero mismatches are mutually incompatible. Consequently, only genotypes $RR$ and $R0$ (the original type) are compatible but $R0$ dies out very early in the simulation leaving all the remaining types incompatible. They can therefore be considered as separate species.

Three stable colonies of genotypes are clearly visible in the gene map from time-step 20000 of a different run (PTAM5-007) from the same simulation, see Figure 5.26 on page 114.

5.3.7.3 PM and AM with initial URZ (PTAM5): conclusions

These results indicate, firstly, that the PM and AM mechanisms can operate successfully together and secondly, that they allow multiple genotypes to persist over long periods. They
also show that PM can create a speciation process in which the original, uniform population $R0$ with one random chromosome (R) paired with a zero chromosome (0) can evolve into two populations: $RR$ and 00 which are mutually incompatible and therefore separate species.

The fact that the population of unmuted genotypes remained largely steady at only 34% & 40% of the population for 18000 time-steps shows that this genotype does not become dominant as it did in the original Penna model. The homozygous $RR$ genotype with two copies of the initial random chromosome (633B) is still well represented (66%) after 20000 time-steps showing that a mutated genotype can persist over time.

The disappearance of the original form $R0$ genotype and the dominance of $RR$ and 00 would seem to be explained as follows. The $R0$ parents can produce gametes $R$ and 0 (with no crossover) which can combine to form zygotes $RR$, 00 or $R0$ all of which may find compatible phenotypes. (In the current model $R0$ and $RR$ produce the same phenotype because all mutations are considered dominant.) However, crossover at a random point can mean that they produce variant gametes which, when combined with other gametes, will be unlikely to find a phenotypically compatible mate. By contrast, the homozygous 00 and $RR$ genotypes will always produce 0 or $R$ gametes, even after crossover, and will therefore be likely to find compatible mates and pass these genes to the next generation. The $RR$ and 00 types therefore have a greater probability of evolutionary success than the original $R0$ type.

This phenomenon is not the same as the Dobzhansky-Müller mechanism and the comparison is discussed in Section 5.3.8.4 on page 118.
5.3.8 De-linked PM and AM with two colonies (DLPA0)

Having established, in the previous simulation PTAM5 (Section 5.3.7 on page 110), that two separate species can arise from a single heterozygous population the next simulation tested the effect of starting with two colonies of different heterozygous genotypes.

In this case, the loci used for PM and AM loci were widely separated on the chromosome unlike the previous simulation in which they were closely interleaved.

Details of the input and results from these simulations are shown in Section C.3.7 on page 213.

5.3.8.1 De-linked PM and AM with two colonies (DLPA0): input

The key input values used in this case are listed below in Table 5.9 on page 115.

<table>
<thead>
<tr>
<th>Run Ref.</th>
<th>Lattice size</th>
<th>DLPA0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rows x Cols</td>
<td>50 x 50</td>
</tr>
<tr>
<td>Recombination</td>
<td>$P_R$</td>
<td>0.60, 1.00</td>
</tr>
<tr>
<td>Mutation</td>
<td>$P_m$</td>
<td>0.00</td>
</tr>
<tr>
<td>PM loci</td>
<td>$L_{pm}$</td>
<td>0xFF00 0000 0000 0000</td>
</tr>
<tr>
<td>PM tolerance</td>
<td>$T_{pm}$</td>
<td>0, 2</td>
</tr>
<tr>
<td>AM loci</td>
<td>$L_{am}$</td>
<td>0x0000 0000 0000 00FF</td>
</tr>
<tr>
<td>AM index</td>
<td>$q_{am}$</td>
<td>4</td>
</tr>
<tr>
<td>Initial A0 (right)</td>
<td>$H_0 &amp; H_1$</td>
<td>0x0000 0000 0000 00FF &amp; 0x0000 0000 0000 00FF</td>
</tr>
<tr>
<td>Initial B0 (left)</td>
<td>$H_0 &amp; H_1$</td>
<td>0xFF00 0000 0000 0000 &amp; 0x0000 0000 0000 0000</td>
</tr>
<tr>
<td>OR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial AA (right)</td>
<td>$H_0 &amp; H_1$</td>
<td>0x0000 0000 0000 00FF &amp; 0x0000 0000 0000 00FF</td>
</tr>
<tr>
<td>Initial BB (left)</td>
<td>$H_0 &amp; H_1$</td>
<td>0xFF00 0000 0000 0000 &amp; 0xFF00 0000 0000 0000</td>
</tr>
<tr>
<td>Loci reported</td>
<td>$L_g$</td>
<td>0x0000 FFFF FFFF 0000</td>
</tr>
</tbody>
</table>

Table 5.9: Summary of assumptions case DLPA0

It can been seen that the loci used in $L_{pm}$ and $L_{am}$ are at opposite ends of the chromosome.

Two initial populations were set up in two halves with uniform colonies covering the left and right halves of the lattice. In the first instance the initial population were heterozygous and second they were homozygous.

The simulation was run for 20000 time-steps and repeated 10 times with the same initial population but with different random seeds which resulted in different sequences of randomised events.
5.3.8.2 De-linked PM and AM with two colonies (DLP A0): results

Because no additional mutations were allowed (i.e. $P_m = 0$) genetic changes could only arise from recombination. When the initial populations were homozygous, recombination generated only gametes identical to the parental chromosomes, and consequently the two species remained entirely separate. This is illustrated in Figure 5.27 on page 116 which shows only the two original genotypes (AA & A0) present after 20000 time-steps even though the colonies have become spatially mixed.

Figure 5.27: Gene-map showing only the two original genotypes after 20000 time-steps (DLP A0-005)

Where the two initial populations were heterozygous and included chromosomes A0 and B0, new zygotes AA, BB, 00 could arise. In these cases the diversity of genotypes generally increased then fell, as in the previous simulation (PTAM5). When $P_R = 0.6$, the peak diversity is lower than when $P_R = 1.0$ because the rate of development of new hybrids was slower with a lower probability of crossover.

For both values of $P_R$, the diversity of genotypes was consistently greater, when phenotype tolerance ($T_{pm}$) was 2, than when it was zero (see Table C.12 on page 214). This is because phenotype selection was being less strictly applied.

In all cases the range of genotypes tended, with increasing time, towards two homozygous types and the zero-zero type.

With crossover probability ($P_R$) set to 0.6 and phenotype tolerance ($T_{pm}$) set to zero (Case DLP A0-000), the unmutilated homozygous genotype (00) tended to dominate soon after 500 time-steps, comprising 40% to 50% of the population while the two homozygous genotypes AA and BB each made up almost all of the other 25% to 30% of the population. These three
homozygous genotypes contained matched chromosomes from each of the initial colonies: \textit{A0} and \textit{B0}.

When the tolerance of phenotype difference ($T_{pm}$ was zero, as in DLPA0-000, after 20000 time-steps the three homozygous genotypes comprised, on average, almost 95\% of the population. The principal three genotypes are shown as red (\textit{AA}), blue (\textit{BB}) and green (00) in Figure 5.28 on page 117, and the detailed percentages are listed in Appendix C.13 on page 217.

![Figure 5.28: Average genotypes populations against time, with strong PM & AM operating ($T_{pm} = 0 \& q_{AM} = 4$). (DLPA0-000)](image)

When some phenotype tolerance was allowed ($T_{pm} = 2$), as in DLPA0-001, the incidence of other hybrids rises to about 15\%. The 23 different genotypes existing in DLPA0-001 after 20000 time-steps are listed in Table C.13 on page 217. Three of these are the homozygous derivatives (\textit{AA}, \textit{BB}, and 00) from the original genotypes \textit{A0} and \textit{B0} but \textit{A0} makes up only 0.108\% and \textit{B0} is absent altogether. Together \textit{AA}, \textit{BB}, and 00 constitute over 94\% of the population. The other 18 genotypes have different mutation patterns, generally at the ends of their chromosomes, as a result of crossovers because no additional mutations were introduced during these simulations ($P_m = 0$).

It is interesting to note that some of the variants in Table C.13 on page 217 are of the form \textit{Ax} where \textit{x}, a less-mutated chromosome, is paired with the original \textit{A} form. In these cases the phenotype will show the expressed genes from \textit{A} because the mutations in \textit{A}, being considered dominant, will be expressed at all the loci, as illustrated in Equation 5.6 on page 118.
\[ \text{Let } A = 01011001 \quad (5.3) \]
\[ \text{Let } x = 01001001 \quad (5.4) \]
\[ A \cup x = 01001101 \quad (5.5) \]
\[ \rightarrow A \cup x = A \quad (5.6) \]

This effect will tend to perpetuate the chromosome \( A \) despite its conjunction with a less-mutated chromosome.

### 5.3.8.3 De-linked PM and AM with two colonies (DLPA0): conclusions

As in the previous simulation PTAM5 (Section 5.3.7 on page 110) these results again demonstrate that the \( A0 \) and \( B0 \) forms of heterozygous genotypes are easily eliminated and are dominated by the homozygous \( AA \) and \( BB \) forms which persist.

This process is similar to the Dobzhansky-Müller mechanism simulations which are recorded in Section 5.4 on page 121. The simulation resembles the Dobzhansky-Müller process because the cross-genotype \( AB \) is not viable, as required by the Dobzhansky-Müller mechanism. \( AB \) does not appear because the mutations for \( A \) and \( B \) are at opposite ends of the chromosome and the phenotypes containing the \( A \) genes will be never match those with the \( B \) genes. This means that phenotypes \( A \) and \( B \) may already be different species which cannot mate.

The apparent speciation process described here is similar to that described above under PTAM5 (Section 5.3.7 on page 110). This phenomenon is not the same as the Dobzhansky-Müller mechanism as discussed below.

### 5.3.8.4 Speciation by phenotype mismatching

Simulations PTAM5 (Section 5.3.7 on page 110) and DLPA0 (Section 5.3.8, above) both started with heterozygous populations: PTAM5 with a single genotype \( R0 \) which evolved to genotypes \( RR \) and \( 00 \), and DLPA0 with equal populations of \( A0 \) and \( B0 \) which evolved to three populations of genotypes \( 00 \), \( AA \) and \( BB \). In both cases these were found to evolve into mutually incompatible species with homozygous genotypes.

These populations were shown to be mutually incompatible because of the requirements of phenotype matching, which represents phenotypic incompatibility between mates. Within the terms of this research, these processes can be considered to be speciation processes. However, in the case of DLPA0 this conclusion is conditional on the phenotypes \( A \) and \( B \) being compatible, otherwise the initial populations of \( A0 \) and \( B0 \) would already have been incompatible species and no new species division would have been observed.
This speciation mechanism could be called “phenotype mismatch speciation” (or PMMS). This mechanism relies on excluding mating between different phenotypes which are so phenotypically different as to be incompatible where these differences are driven by a pair of alleles one dominant and the other recessive.

Two immediate questions may be raised about PMMS. Firstly, is it a variant of the Dobzhansky-Müller mechanism; and secondly, does it represent a process present in nature?

The Dobzhansky-Müller mechanism, which was described in detail in Section 2.3.2 on page 31, concerns changes at two genetic loci, designated $a$ and $b$. In a large population of genotype $aabb$ small colonies might mutate to $Aabb$ and $aaBb$. These genotypes could then evolve to $AAbb$ and $aaBB$ which are incompatible because the coincidence of $A$ and $B$ is assumed to be fatal to the zygote. By comparison, PMMS depends on the rejection of mates, prior to mating, which is based on characteristics driven by phenotypically expressed genes driven by alleles which are either dominant and recessive.

Dobzhansky-Müller relies on gene coincidences which are fatal in the zygote whereas PMMS depends on the structural incompatibility of phenotypically expressed genetic characteristics prior to mating. The two mechanisms are therefore substantially different in their genetic processes.

The question of natural occurrences of PMMS is more difficult to address because one’s knowledge of the whole natural world is necessarily incomplete. However, in order for the process to occur there would need to be a dominant and recessive allele at a locus which, when phenotypically expressed, are incompatible with each other to the extent that they are reproductively infertile. If the two alleles were $P$ and $p$, then the PMMS simulations imply that a population of genotype $Pp$ would evolve into two mutually incompatible, homozygous genotypes $PP$ and $pp$.

Because, in these circumstances, evolutionary transition into separate species seems (according to these simulations) to be inevitable, it might be difficult to identify natural occurrences because these would already have completed their speciation. However, it is possible that plants which exhibit specialised flower forms, such as the *Mimulus*, described in Section 2.3.3 on page 32, might have become distinct in this way. These plants attract different pollinators which need different flower structures which only cross-pollinate similar flowers. The resulting different varieties of *Mimulus* are not separate species because they are not genetically incompatible because they can be artificially crossed successfully but they remain separate because of the preferences exhibited by their specialised pollinators.

It would be interesting to investigate how the two forms of *Mimulus* originated but that is beyond the scope of the current research. Wu et al. [2008] give a comprehensive account of the genetics of the genus *Mimulus* and its usefulness in genomic studies.
5.3.9 Unconstrained diversity

For completeness, some further simulations were run with no constraint (PM or AM) on the choice of mate or the development of zygotes. These simulations were run with $P_R$ set to 0.2, 0.4 and 0.8, and with the standard three initial populations: URC, URZ and RZ.

As might be expected, in all cases this resulted in the diversity of genotypes rising continuously and tending towards the population size (2500) with almost as many different genotypes as individuals. The cases which started from a single uniform genotype (URC & URZ) increased towards 2500. Alternatively when starting with a randomised population (in the RZ case) in which there were 2500 different genotypes, the diversity decreased slightly reflecting a general population decline. The rate of change was higher where the probability of crossover ($P_R$) was higher.

These trends are illustrated in Figures C.22 to C.24 in Appendix C.3.8 on page 218.

These results demonstrate that, if unconstrained by any selection mechanisms, diversity will tend to increase to one genotype per individual. They are not unexpected results and do not materially contribute to the search for speciation but they give some reassurance that the model is working as expected.
5.4 Simulations with the Dobzhansky-Müller mechanism

The simulations described in this section investigate the behaviour of the Dobzhansky-Müller (DM) mechanism\(^4\) which proposes that separate species can be formed when there are at least two genetic changes in the genome and an intermediate step is non-viable. The Dobzhansky-Müller mechanism was described more fully in Section 2.3.2 on page 31.

It was possible to undertake some simulations of the Dobzhansky-Müller (DM) mechanism using phenotype matching and assortative mating in various degrees and with the Penna ageing process turned off. The range of simulations of the DM mechanisms is summarised in the following sections.

In simulation DM04 (Section 5.4.1) the basic form of the mechanism was demonstrated. Then simulation DM07 (Section 5.4.2) tested its sensitivity to the introduction of assortative mating with various values for the AM index \((q_{am})\) and the mating range \((R_m)\). In DM08 (Section 5.4.3) further tests investigated the effect of the separation of the key Dobzhansky-Müller loci on the chromosome. Simulations DM09 & DM10 (Section 5.4.4) briefly looked at the extreme cases with no exclusion of the \((AB)\) phenotype and no addition of mutations, respectively.

The initial populations in DM11 (Section 5.4.5) contains only genotypes \(Aabb\) & \(aaBb\) but allows no additional mutations. Simulation DM12 (Section 5.4.6) attempts to simulate the DM mechanism starting from small isolated colonies of genotypes \(Aabb\) & \(aaBb\) within a general population of \(aabb\) genotype, but shows that these colonies generally do not survive.

5.4.1 Initial Dobzhansky-Müller simulation (DM04)

In this simulation the simplest case was tested looking at changes at only the lowest two loci of the chromosome, bits 0 and 1. The changes involved the binary possibilities \(...00, ...01, ...10\) and \(...11\). These were taken as representing the DM alleles: \(...ab, ...aB, ...Ab\) and \(...AB\), where the capital letters indicate mutated alleles. The ancestor population in the DM mechanism, \(aabb\), was therefore represented with \(...00\) in both chromosomes.

The non-viable phenotype combination, \(AB\), could arise from any genotype including both \(A\) and \(B\), these being taken as the dominant alleles. Any phenotype expressing \(...11\) therefore had to be destroyed.

In order to simplify the number of possible outcomes which might arise from random mutations at other loci, all the other loci on the chromosomes were assumed to be already mutated (1s). Because mutations are only additive in this model, there could be no further changes at these loci.

---

\(^4\)In the current text the term ‘Dobzhansky-Müller mechanism’ (or ‘DM mechanism’) is used to refer to the evolutionary process generally called ‘The Dobzhansky-Müller model’ in the literature. This has been done to avoid confusion with the term model used in this thesis to refer to computer models.
Fuller details of the input and output data are given in Appendix C.4.1 on page 222.

5.4.1.1 Initial Dobzhansky-Müller simulation (DM04): input

The first tests were to assess the impact on the DM mechanism of phenotype matching (PM) and assortative mating (AM). These mechanisms were therefore turned on and off in four separate cases shown in Table 5.10 on page 122.

<table>
<thead>
<tr>
<th>Run ref</th>
<th>PM</th>
<th>AM</th>
</tr>
</thead>
<tbody>
<tr>
<td>DM04-000</td>
<td>Off</td>
<td>Off</td>
</tr>
<tr>
<td>DM04-001</td>
<td>Off</td>
<td>On</td>
</tr>
<tr>
<td>DM04-002</td>
<td>On</td>
<td>Off</td>
</tr>
<tr>
<td>DM04-003</td>
<td>On</td>
<td>On</td>
</tr>
</tbody>
</table>

Table 5.10: DM04 Use of Phenotype matching and Assortative mating

A high value, 0.80, for recombination probability ($P_R$) was used but this was subsequently shown to be irrelevant. The probability of mutation during meiosis ($P_m$) was set to 1.0. A high mating range ($R_M = 10$) was set in order to give plants an opportunity to reach similar plants for mating. When assortative mating was used an AM index ($q_{am}$) of 8 was used to emphasise the impact of that mechanism. Some tolerance was allowed in the application of phenotype matching: matching was tested at the $A$ & $B$ loci and but a matching tolerance of 1 was allowed so that mating could occur with a match at only one of these loci.

Later tests (DM07, see 5.4.2) investigated the influence of changing the AM index and the mating range.

5.4.1.2 Initial Dobzhansky-Müller simulation (DM04): results

In general, the results show that the unmutated form $aabb$ is progressively replaced by the mutated forms $AAbb$ and $aaBB$ as predicted by the Dobzhansky-Müller model. Some results are reproduced in this section for easy reference with more detailed results shown in Appendix C.4.1.2 on page 222.

With AM and PM switched off, in case DM04-000 (Figure 5.29 on page 123), the population shifts only slowly to the fully mutated forms $AAbb$ and $aaBB$ and, even after 100000 time-steps, there is still a small part of the population (0.2%) with the original genotype $aabb$. It can also be seen from the genotype mapping in Figure C.26 on page 225 that no coherent colonies of each genotype are forming: the population remains randomly spread until one type finally dominates.

In cases DM04-001 & DM04-003, where AM is applied there is a marked acceleration in the genetic change. With AM present without PM, in case DM04-001 (Figure 5.30 on page 123),
the \textit{aabb} genotype has substantially disappeared after 30000 time-steps and completely gone by 60000, and the genotype maps (Figure C.28 on page 227) show the formation of flocculating colonies of the same genotype. However, by 100000 time-steps, one type (\textit{aaBB}) seems to be starting to dominate, tending on average towards 60\% of the population.

With PM switched on, in case DM04-002 (Figure 5.31 on page 124), the population moves slightly more slowly, than with no PM (DM04-000), to the mutated forms and after 100000 time-steps 6.6\% of the population still has the original \textit{aabb} genotype. PM, as applied here allowing one genetic difference at two loci, seems to have limited effect on the genetic change in the population.

With PM and AM working together in case DM04-003 (Figure 5.32 on page 124), the \textit{aabb} genotype and the intermediate forms \textit{Aabb} and \textit{aaBb} are eliminated completely in 40000 time-
Figure 5.31: Slow population shift with PM only. (DM04-002)

steps and the populations of $AAbb$ and $aaBB$ remain stable and approximately evenly divided for the next 60000 time-steps. The genotype maps of this case also show a tendency to regional stability.

Figure 5.32: Rapid population shift with AM and PM (DM04-003)
5.4.1.3 Initial Dobzhansky-Müller simulation (DM04): conclusions

In this simplified case, the Dobzhansky-Müller model appears to allow successful sympatric speciation to occur, but with the continual insertion of mutations ($P_m = 1$).

It is also apparent that assortative mating contributes significantly to the acceleration of the process and the establishment of cohesive colonies.

It will be interesting to investigate, firstly, the sensitivity of this process to changes in application of AM and PM, and secondly, whether the arrival of a predator might provide the catalyst for the observed population changes.

5.4.2 Dobzhansky-Müller sensitivity simulation (DM07)

In order to investigate the influence of mating range and the assortative mating index on the Dobzhansky-Müller scenario a further series of simulations was run.

At this stage it should be noted that changes in the probability of crossover ($P_R$) are not expected to influence the process because in this simplified situation crossovers have no effect for the following reason. The initial population contains only the genotype $aabb$ to which may be added, after random mutation, the genotypes $Aabb$ and $aaBb$, but genotype $AaBb$ is specifically excluded. The resulting genotypes only have genes of interest at two loci ($a$ and $b$) and therefore in a diploid configuration crossing-over between these two loci cannot produce any additional type of gamete. It should also be noted that, for simplicity, all the other genes are initially set to mutated (1) so that no more changes can occur at these loci.

The gene positions and possible gametes from the three permitted genotypes are shown in Table 5.11.

<table>
<thead>
<tr>
<th>Parent genotype</th>
<th>Chromosome genes</th>
<th>Possible gametes</th>
</tr>
</thead>
<tbody>
<tr>
<td>$aabb$</td>
<td>a..b</td>
<td>ab</td>
</tr>
<tr>
<td></td>
<td>a..b</td>
<td>ab</td>
</tr>
<tr>
<td>$Aabb$</td>
<td>A..b</td>
<td>Ab</td>
</tr>
<tr>
<td></td>
<td>a..b</td>
<td>ab</td>
</tr>
<tr>
<td>$aaBb$</td>
<td>a..B</td>
<td>aB</td>
</tr>
<tr>
<td></td>
<td>a..b</td>
<td>ab</td>
</tr>
</tbody>
</table>

Table 5.11: DM07 Possible gametes

It can be seen that the range of possible gametes will remain the same whether or not crossover occurs between loci $a$ and $b$. 

125
5.4.2.1 Dobzhansky-Müller sensitivity simulation (DM07): input

The impact of assortative mating was varied with the AM index \((q_{am})\) set at 0, 2, 4 & 8 and mating range \((R_m)\) was variously set at 1, 4 & 8. These scenarios were all run for 100000 time-steps and repeated 20 times. The full input data is shown in Appendix C.4.2.1 on page 232.

5.4.2.2 Dobzhansky-Müller sensitivity simulation (DM07): results

Details of the output data are shown in Appendix C.4.2.2 on page 232 but Table 5.12 summarises the inputs and the key outcome for each run.

<table>
<thead>
<tr>
<th>Run ref</th>
<th>Ass. mating index ((q_{am}))</th>
<th>Mating range ((R_m))</th>
<th>Final pop. % (aabb)</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>DM07-000</td>
<td>0</td>
<td>1</td>
<td>15.0%</td>
<td>(aabb) persists; (AAbb) &amp; (aaBB) grow slowly</td>
</tr>
<tr>
<td>DM07-001</td>
<td>0</td>
<td>4</td>
<td>15.5%</td>
<td>(aabb) persists; (AAbb) &amp; (aaBB) grow slowly</td>
</tr>
<tr>
<td>DM07-002</td>
<td>0</td>
<td>8</td>
<td>7.1%</td>
<td>(aabb) persists; (AAbb) &amp; (aaBB) grow slowly</td>
</tr>
<tr>
<td>DM07-003</td>
<td>2</td>
<td>1</td>
<td>0.0%</td>
<td>Only (AAbb) &amp; (aaBB) after (t=50k)</td>
</tr>
<tr>
<td>DM07-004</td>
<td>2</td>
<td>4</td>
<td>84.5%</td>
<td>(aabb) persists with very little other development</td>
</tr>
<tr>
<td>DM07-005</td>
<td>2</td>
<td>8</td>
<td>89.8%</td>
<td>(aabb) persists with very little other development</td>
</tr>
<tr>
<td>DM07-006</td>
<td>4</td>
<td>1</td>
<td>0.0%</td>
<td>Only (AAbb) &amp; (aaBB) after (t=45k)</td>
</tr>
<tr>
<td>DM07-007</td>
<td>4</td>
<td>4</td>
<td>0.0%</td>
<td>Only (AAbb) &amp; (aaBB) after (t=80k)</td>
</tr>
<tr>
<td>DM07-008</td>
<td>4</td>
<td>8</td>
<td>76.0%</td>
<td>(aabb) persists with very little other development</td>
</tr>
<tr>
<td>DM07-009</td>
<td>8</td>
<td>1</td>
<td>0.0%</td>
<td>Only (AAbb) &amp; (aaBB) after (t=50k)</td>
</tr>
<tr>
<td>DM07-010</td>
<td>8</td>
<td>4</td>
<td>0.0%</td>
<td>Only (AAbb) &amp; (aaBB) after (t=50k)</td>
</tr>
<tr>
<td>DM07-011</td>
<td>8</td>
<td>8</td>
<td>0.0%</td>
<td>Only (AAbb) &amp; (aaBB) after (t=45k)</td>
</tr>
</tbody>
</table>

Table 5.12: DM07 Influence of AM index \((q_{AM})\) and mating range \((R_m)\) on \(aabb\) population at time 100000

5.4.2.3 Dobzhansky-Müller sensitivity simulation (DM07): conclusions

The mating range seems to work against assortative mating by offering a larger group of mates and thereby increasing the probability that an ‘unsuitable’ mate will be selected. Raising the assortative mating index \((q_{am})\) compensates for this by applying a more discerning selection of
a mate, but with a large mating range, such as 8 which includes 63 possible mates, the index must be raised to 8 in order to achieve the full conversion of the population to the two new species $AAbb$ and $aaBB$.

### 5.4.3 Dobzhansky-Müller de-linking simulation (DM08)

This simulation was devised to test the effect of separating on the chromosome (de-linking) the key DM loci which represent $A$ and $B$. Other inputs similar to the previous simulations were used.

#### 5.4.3.1 Dobzhansky-Müller de-linking simulation (DM08): input

Previously bit positions 1 & 0 had been used to represent gene loci $A$ & $B$ but in these simulations bits 47 and 16 were used. The loci of $A$ & $B$ were therefore defined by the two bits in $0x0000800000010000$. These were chosen to be some distance from the chromosome ends, which limits the effects of crossovers, and also substantially separates them from each other.

With this de-linking of $A$ & $B$, the influence of increasing crossover probability was tested by setting the probability of crossover ($P_R$) to 0.2, 0.4 and 0.8. Crossover was not expected to influence the outcome as explained in 5.4.2 above, but any influence would be more pronounced by de-linking the key loci.

Having established in simulation DM07 that a small mating range enhances the influence of AM, here the mating range was set to 1. The AM index ($q_{am}$) was set to 4 and 8 in order to provide a strong AM influence as used in simulation DM07.

The six input cases are listed in Table 5.13 on page 128 and full input details are given in Appendix C.4.3.1 on page 258.

#### 5.4.3.2 Dobzhansky-Müller de-linking simulation (DM08): results

Detailed results are shown in the appendix: Figure C.57 on page 260 to Figure C.68 on page 271 but the key outcomes are listed in Table 5.13. In all cases the original $aabb$ genotype disappeared at the times listed.

The $AAbb$ & $aaBB$ genotypes were approximately equally represented by the end of 100000 time-steps but the average number of genotypes was approaching 1.5 indicating that in a substantial proportion of the 20 reruns, only one genotype persisted until time 100000. As a typical example, the average populations and number of genotypes in case DM08-000, with crossover probability of 0.20 and AM index 4, are shown in Figure 5.33 on page 128.
<table>
<thead>
<tr>
<th>Run ref</th>
<th>Crossover $P_R$</th>
<th>AM index $q_{am}$</th>
<th>Approx time to only AAbb &amp; aaBB at Time 100 000</th>
<th>Av No of genotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>DM08-000</td>
<td>0.20</td>
<td>4</td>
<td>40 000</td>
<td>1.55</td>
</tr>
<tr>
<td>DM08-001</td>
<td>0.20</td>
<td>8</td>
<td>45 000</td>
<td>1.55</td>
</tr>
<tr>
<td>DM08-002</td>
<td>0.40</td>
<td>4</td>
<td>40 000</td>
<td>1.60</td>
</tr>
<tr>
<td>DM08-003</td>
<td>0.40</td>
<td>8</td>
<td>50 000</td>
<td>1.65</td>
</tr>
<tr>
<td>DM08-004</td>
<td>0.80</td>
<td>4</td>
<td>40 000</td>
<td>1.75</td>
</tr>
<tr>
<td>DM08-005</td>
<td>0.80</td>
<td>8</td>
<td>50 000</td>
<td>1.60</td>
</tr>
</tbody>
</table>

Table 5.13: DM08 Assumptions for testing separated loci for Aa and Bb

Figure 5.33: Population shift to two new species: $P_R = 0.2$ & $q_{am} = 4$ (DM08-000)
5.4.3.3 Dobzhansky-Müller de-linking simulation (DM08): conclusions

The longterm results appear to be largely consistent with the earlier simulations and show insensitive to delinking and to changes in the probability of crossover ($P_r$) and AM index ($q_{am}$). However an increase in the probability of crossover, while not affecting the ultimate outcome, seems to accelerate the rate of change.

These simulations confirm that the Dobzhansky-Müller mechanism appears to operate with separated (de-linked) loci as well as with closely linked loci.

5.4.4 Dobzhansky-Müller special cases (DM09 & DM10)

For completeness the model was run with two special cases, using similar inputs to DM08, with separated loci for the genes $A$ & $B$.

5.4.4.1 Dobzhansky-Müller special cases (DM09 & DM10): input

In the first case (DM09) the intermediate $AB$ phenotypes were not excluded as required in the standard Dobzhansky-Müller mechanism. In the second case (DM10), starting with a whole population of $aabb$, no additional mutations were allowed (i.e. $P_m = 0$).

Details of these simulations can be seen in Appendix C.4.4 on page 272 and Appendix C.4.5 on page 276.

5.4.4.2 Dobzhansky-Müller special cases (DM09 & DM10): results

As expected in DM09, where the $AB$ phenotype is not excluded, the genotype $AABB$ dominated the population within 50000 time-steps. This genotype is the fully mutated form of the original $aabb$ genotype and since mutations were added but never removed it is inevitable that this form will eventually dominate.

The DM10 case is trivial because without the addition of any mutations there can be no evolutionary change and the population is genetically static with only the $aabb$ genotype throughout the simulation. Crossovers have no influence in this case because the $aabb$ form is homozygous so that recombinations of their chromosomes are indistinguishable from the original.

5.4.4.3 Dobzhansky-Müller special cases (DM09 & DM10): conclusions

Failure to exclude the $AB$ phenotype means that the Dobzhansky-Müller mechanism does not operate and evolution to two incompatible genotypes does not occur.

Failure to add any mutations inevitably means that the population remains unaltered over time.
These results demonstrate that the model is working as expected even if the results of these simulations are unremarkable.

5.4.5 Dobzhansky-Müller starting from $Aabb$ and $aaBb$ (DM11)

In these simulations the initial population was divided equally between $Aabb$ and $aaBb$ genotypes. These are each one mutation different from the basic $aabb$ genotype and it is assumed that these mutations are already established in adjacent colonies. This simulation was run without excluding the $AB$ phenotype but using assortative mating.

It was intended to investigate whether the Dobzhansky-Müller mechanism could be successful with large numbers of independent $A$ & $B$ alleles present from the start but without excluding the $AB$ phenotype.

Details of this simulation can be seen in Appendix C.4.6 on page 280.

5.4.5.1 Dobzhansky-Müller starting from $Aabb$ and $aaBb$ (DM11): input

The initial population was divided into two contiguous areas of half the lattice, populated with $Aabb$ and $aaBb$ genotypes respectively. No addition of mutations was allowed during the simulation. The probability of crossover ($P_R$) was set to 0.80. The simulation was run with assortative mating inactive ($q_{am} = 0$) and then active ($q_{am} = 4$), as summarised in Table 5.14.

Most significantly, in both cases there was no exclusion of the $AB$ phenotype so the Dobzhansky-Müller process was not expected to operate properly.

The two cases were rerun 20 times and the results averaged.

<table>
<thead>
<tr>
<th>Run ref</th>
<th>Scenario</th>
<th>Crossover</th>
<th>AM index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$P_R$</td>
<td>$q_{am}$</td>
</tr>
<tr>
<td>DM11-000</td>
<td>Without AM</td>
<td>0.80</td>
<td>0</td>
</tr>
<tr>
<td>DM11-001</td>
<td>With AM</td>
<td>0.80</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 5.14: DM11 Assumptions for testing AM, without excluding the $AB$ phenotype

5.4.5.2 Dobzhansky-Müller starting from $Aabb$ and $aaBb$ (DM11): results

When AM was not active (case DM11-000) the genotypes stabilised with the unmutated $aabb$ form constituting just over 40% of the population. After 100000 time-steps about 7% of the population have $AB$ alleles, this proportion having declined from a maximum of just over 10% at about 50000 time-steps. The $AB$ alleles are present in the genotypes $AaBb$, $AaBB$, $AABb$ and $AABB$ shown in Figure 5.34 on page 131.
However, when AM was applied (case DM11-001) the incidence of the $AB$ phenotypes rises constantly and it seems that $AABB$ would ultimately dominate the population, as shown in Figure 5.35 on page 131. The unmutated $aabb$ genotype has completely disappeared by 70000 time-steps.

Figure 5.35: Population shift with AM applied ($q_{am} = 4$) and no $AB$ exclusion (DM11-001)
5.4.5.3 Dobzhansky-Müller starting from $Aabb$ and $aaBb$ (DM11): conclusions

This simulation confirms that the Dobzhansky-Müller process does not function fully when there is no exclusion of the $AB$ phenotypes irrespective of whether AM is applied.

However, when AM is not applied the $AB$ phenotypes seem to have difficulty becoming established whereas with AM active $AABB$, the fully mutated genotype, soon dominates the population. The reason for this difference in response is unclear but it may be that without AM the initially dominant alleles are able to persist but with AM active the $A$ & $B$ alleles become established because they are mutually attracted through the AM selection.

5.4.6 Dobzhansky-Müller with initial small areas of $Aabb$ and $aaBb$ (DM12)

In the earlier Dobzhansky-Müller simulations the initial situation either required the original population of genotype $aabb$ to be subject to continual mutation, as in DM04 and DM07 (Section 5.4.1 on page 121 & Section 5.4.2 on page 125) or a large population of the ‘mutants’ $Aabb$ and $aaBb$ genotypes were required at the start, as in DM11.

Simulation DM12 was therefore conceived to investigate whether small instances of mutation could develop into stable colonies. This was set up with two small (5x5) colonies of $Aabb$ and $aaBb$ genotypes in opposite corners of the lattice which was generally populated with the unmutated $aabb$ genotype. This starting position was assumed to represent the situation immediately after random mutations had generated the first examples of the mutants. The separation of the mutant colonies at opposite corners of the lattice was chosen on the assumption that the mutants would not have arisen in close proximity, but other starting configurations would be equally valid.

Details of this simulation can be seen in Appendix C.4.7 on page 286.

5.4.6.1 Dobzhansky-Müller with initial small areas of $Aabb$ and $aaBb$ (DM12): input

The simulation was run with the usual Dobzhansky-Müller constraint that $AB$ phenotypes are non-viable, but with varying rates for the incidence of new mutations ($P_m = 0.0, 0.1, 0.2, 0.3, 0.4, 0.5$). Each scenario was run for 20000 time-steps which was found to be sufficient time to see whether stable colonies were developing or the mutations had died out.
5.4.6.2 Dobzhansky-Müller with initial small areas of $AAbb$ and $aaBb$ (DM12): results

Where no additional mutations are introduced ($P_m = 0$) in case DM12-000, the small colonies of mutants generally reduce in size but their homozygous derivatives $AAbb$ & $aaBB$ genotypes tend to persist in small numbers, as shown in Table 5.15. Here the 5x5 initial colonies each amounts to 1% of the overall 50x50 lattice. It can be seen that the $AAbb$ & $aaBB$ genotypes have not reached 1% of the population after 20000 time-steps.

Table 5.15: Population shift with no additional mutations ($P_m = 0$) (DM12-000)

<table>
<thead>
<tr>
<th>Average for all runs of percentage of maximum population for each genotype</th>
<th>Time</th>
<th>0</th>
<th>5000</th>
<th>10000</th>
<th>15000</th>
<th>20000</th>
<th>50000</th>
<th>100000</th>
<th>150000</th>
<th>200000</th>
<th>500000</th>
</tr>
</thead>
<tbody>
<tr>
<td>$aabb$</td>
<td>95.5%</td>
<td>96.6%</td>
<td>95.5%</td>
<td>95.5%</td>
<td>96.6%</td>
<td>95.5%</td>
<td>96.6%</td>
<td>95.5%</td>
<td>96.6%</td>
<td>95.5%</td>
<td>96.6%</td>
</tr>
<tr>
<td>$aaBB$</td>
<td>5.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>AAbb</td>
<td>0.5%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Aabb</td>
<td>5.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>$AABB$</td>
<td>0.5%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>$aaBS$</td>
<td>5.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>$aABS$</td>
<td>0.5%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>$AB$</td>
<td>5.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>$AB$</td>
<td>0.5%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

With the addition of more mutations the mutant population develops, developing a larger population as the mutation rate increases. Tables 5.16 and 5.17 illustrate the outcome with increasing probabilities of mutation, $P_m = 0.1$ and $P_m = 0.5$, respectively.

Table 5.16: Population shift with 10% probability of mutations ($P_m = 0.1$) (DM12-001)

<table>
<thead>
<tr>
<th>Average for all runs of percentage of maximum population for each genotype</th>
<th>Time</th>
<th>0</th>
<th>1000</th>
<th>2000</th>
<th>5000</th>
<th>10000</th>
<th>15000</th>
<th>20000</th>
<th>50000</th>
<th>100000</th>
<th>150000</th>
<th>200000</th>
</tr>
</thead>
<tbody>
<tr>
<td>$aabb$</td>
<td>99.9%</td>
<td>98.8%</td>
<td>97.5%</td>
<td>96.0%</td>
<td>94.6%</td>
<td>92.1%</td>
<td>90.6%</td>
<td>90.6%</td>
<td>90.6%</td>
<td>90.6%</td>
<td>90.6%</td>
<td>90.6%</td>
</tr>
<tr>
<td>$aaBB$</td>
<td>1.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>AAbb</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Aabb</td>
<td>1.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>$AABB$</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>$aaBS$</td>
<td>1.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>$aABS$</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>$AB$</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>$AB$</td>
<td>1.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

Table 5.17: Population shift with 50% probability of mutations ($P_m = 0.5$) (DM12-005)

In all three cases it can be seen that coincidences of $A$ & $B$ (shown in red) have not been allowed to appear, as required under the Dobzhansky-Müller scenario.
5.4.6.3 Dobzhansky-Müller with initial small areas of $Aabb$ and $aaBb$ (DM12): conclusions

This series of simulations seems to show that small mutant populations fail to grow in size because they are overwhelmed by the initial $aabb$ genotypes but with a higher rate of continual mutation, the mutant population does become established. However, this indicates that the growth in the number of mutants is probably the result of additional mutations occurring after the start of the simulation and does not necessarily derive from the initial mutant colonies. This implies that development from small colonies is not sustainable within a large unmutated population and some other intervention is needed to find circumstances in which small mutant colonies might become established.

Subsequent simulations show how small initial mutant colonies might be preserved by the intervention of a predator.
5.5 Simulations of Dobzhansky-Müller with predators

During the investigation of the Dobzhansky-Müller mechanism two possible topics for further investigation became apparent.

- Could a predator be involved in the exclusion of the $AB$ phenotypes
  or, alternatively,

- could a predator be helpful in allowing a colony of mutants to become established?

This series of simulations investigates the application of predators in these ways to the Dobzhansky-Müller mechanism.

The mechanisms of predation in the *Sympatria* model is explained in detail in Section 4.7.3 on page 72. In summary, predators are assumed to be attracted or deterred by certain phenotypic characteristics due to expressed genes $E_{Pa}$ and $E_{Pd}$ at loci $L_{Pa}$ and $L_{Pd}$ respectively, all of which are specified in the input. If the attraction characteristics are expressed and the deterrent are not expressed then the predation rate, $D_P$, is added to the general environmental death rate ($D_0$). As in other parts of the model, the genes expressing characteristics in the phenotype are taken as the ‘OR’ function of the genes in its two chromosomes.

5.5.1 Predation as the $AB$ excluder (PR00)

The earlier simulations of the Dobzhansky-Müller mechanism used an explicit ‘fatal phenotype’ function to exclude the $AB$ phenotype. In the next set of simulations (PR00) the exclusion of $AB$ was implemented by a predator which provided an enhanced death rate for organisms of that phenotype.

Fuller details of the input and output data are given in Appendix C.5.1 on page 300.

5.5.1.1 Predation as the $AB$ excluder (PR00): input

The input assumptions are similar to those in the previous Dobzhansky-Müller simulations except that there is no specific exclusion of the $AB$ phenotypes at their inception. Instead a predator is set up which is attracted to the $AB$ phenotype.

In the first two cases (PR00-000 & -001) the $AB$ phenotypes attract a predation rate ($D_P$) of 50% and in PR00-002 & -003 this is increased to 100%, which means that the predator will kill 100% of the prey which exhibit the target phenotype characteristics. In each pair of cases the assortative mating index ($q_{am}$) is set to 0 (for no AM), and then 4 (for strong AM).

In all cases the general death rate ($D_0$) was left at the default value of 1% and consequently the predation rates are likely to dominate the scenario.
The initial population was equally spatially divided between the Aabb and aaBb genotypes and no additional mutations or crossovers were allowed, \( P_R = 0 \) & \( P_m = 0 \).

As discussed in Section 4.7.3.1 on page 72, the total elimination of the prey by a predator might result in the death of the predators, but total elimination of the prey is not an invalid assumption if it only occurs once in the evolutionary process.

5.5.1.2 Predation as the AB excluder (PR00): results

In case (PR00-000) with only 50% predation rate but no AM, the Dobzhansky-Müller mechanism is not effective and the unmuted genotype aabb became dominant, as shown in the population chart Figure 5.36 on page 136. After 100000 time-steps genotype aabb still comprises 70% of the population and is progressively excluding the mutated forms.

However, the AB phenotypes are completely absent throughout the simulation which indicates that the predator is effective.

In the case (PR00-001) with a predation rate of 50% again, but with AM \( q_{am} = 4 \), speciation is successfully achieved and by 80000 time-steps only the AAbb and aaBB genotypes are present, as shown in Figure 5.37 on page 137.

Again, no AB phenotypes are seen during the simulation because of the predator.

Simulations PR00-002 & -003, in which the predation rate is set to 100% \( (D_P = 1.0) \), similar effects are observed but with a very slightly increased rate of change.
5.5.1.3 Predation as the AB excluder (PR00): conclusions

These simulations demonstrate that a predator can act to exclude the AB phenotypes in the Dobzhansky-Müller mechanism, particularly when assortative mating is active.

However, the pre-existing population used in these cases was wholly divided between the heterozygous genotypes $Aabb$ and $aaBb$. The simulations therefore represent only part of the Dobzhansky-Müller mechanism and do not start from the completely unmutated $aabb$ genotype as required for the full Dobzhansky-Müller process.

The cases PR00-000 & PR00-002 without assortative mating, speciation was unsuccessful and the $aabb$ genotype appears to be able to dominate the population. This probably results of crosses from the mutated form and the pre-existing $ab$ chromosome which was present in both the initial heterozygous genotypes $Aabb$ and $aaBb$. The prevalence of the $aabb$ genotype seems to limit the development of colonies of the new mutant forms $AAbb$ and $aaBB$.

It was therefore decided to try using a predator to deplete the $aabb$ population in order to allow the mutant genotypes to flourish, instead of using the predator to exclude the $AB$ phenotypes.

5.5.2 Predation to deplete the $aabb$ forms (PR01)

In the first simulations of the Dobzhansky-Müller mechanism (see 5.4) a high rate of mutation was used which continually converted $aabb$ genotypes to $Aabb$ or $aaBb$. From the subsequent DM12 simulations (see Section 5.4.6 on page 132), it appears that this continual insertion of mutations is necessary to avoid the replacement of the mutant genotypes with the unmutated $aabb$ genotype.
An alternative, used in simulations DM11 and PR00, started with the population already divided between $Aabb$ and $aaBb$ genotypes but this scenario seems inadequate because it assumes that a significant evolutionary step away from the $aabb$ genotypes has already occurred. Both these methods seem to be unnatural constructs and therefore a different approach was sought.

In the next simulations the initial population consisted of almost entirely $aabb$ genotypes but with very small (2x2), nascent colonies of $Aabb$ & $aaBb$ genotypes at opposite corners of the lattice. This is similar to the earlier simulation DM12 (Section 5.4.6 on page 132) in which 5x5 colonies were used, but in order to simulate the development from very small instances of mutation the initial colony size has been reduced to 2x2. No additional mutation was allowed during the simulation.

In the DM12 simulations it was found that the small mutant colonies, were overrun by the $aabb$ genotype. To counteract this a predator is now introduced to deplete the $aabb$ population as if a disease or some other threat has just arrived in the region.

During this process the $AB$ phenotypes are excluded as zygotes simulating some developmental deficiency of that phenotype. This is the same mechanism as used in the first Dobzhansky-Müller simulation DM04 (Section 5.4.1 on page 121).

Fuller details of the input and output data for the PR01 simulation are given in Appendix C.5.2 on page 310.

5.5.2.1 Predation to deplete the $aabb$ forms (PR01): input

Small (2x2) colonies of $Aabb$ & $aaBb$ genotypes are initially inserted at opposite corners of the lattice which is otherwise filled with $aabb$ genotypes. This might represent the moment when mutations for $Aabb$ & $aaBb$ have just arisen. The simultaneous occurrence of these mutations is not particularly significant because the development of the colonies of $Aabb$ and $aaBb$ are effectively independent in the early stages of the process so that it would make little difference if these colonies were initiated at different times.

Several cases were run with predation rates ($D_P$) of 0%, 5% & 10%, attacking the $ab$ phenotype which can only arise from the $aabb$ genotype.

Assortative mating is active, with a strong index of $q_{am} = 8$, focused on the phenotypic characteristics expressed at the loci of $a$ and $b$ so that all phenotypes would be attracted to their own kind.

These simulations were run for only 500 time-steps which was found to be sufficient time to see whether a substantial population of mutants had started to develop.
5.5.2.2 Predation to deplete the $aabb$ forms (PR01): results

The genotype maps in Figure 5.38 on page 140 (for two separate, sample runs) show how an increase in the predation rate ($D_P$) allows the small initial mutant colonies of $Aabb$ and $aaBb$ to develop. In all cases the small mutant colonies can be seen at the corners at time zero.

In case PR001-000, with no predation ($D_P = 0$), the mutant colonies died out (or almost died out) after 500 time-steps. In the subsequent cases (PR001-001 & -002) as the predation rate is increased the mutant colonies are able to become established. With $D_P = 10\%$ (PR001-002), the two mutant colonies have almost filled the lattice in 500 time-steps.
### Table 5.38: Gene Map Development at Various Predation Rates

<table>
<thead>
<tr>
<th>Case:</th>
<th>[All]</th>
<th>PR01-000</th>
<th>PR01-001</th>
<th>PR01-002</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time:</td>
<td>0</td>
<td>500</td>
<td>500</td>
<td>500</td>
</tr>
<tr>
<td>Predation ($D_P$):</td>
<td>0%</td>
<td>5%</td>
<td>10%</td>
<td></td>
</tr>
</tbody>
</table>

![Gene Maps](image)

**Figure 5.38:** Two examples of gene maps showing colony development at 500 time-steps with various predation rates ($D_P$) depleting the $aabb$ genotype. Genotypes are colour-coded. (PR01)
5.5.2.3 Predation to deplete the $aabb$ forms (PR01): conclusions

These simulations demonstrate that a predator, with $D_P \geq 5\%$, acting on the general population can facilitate the establishment of new mutant colonies which develop from small localised instances of new mutation.

It should be noted that the predator used in this example is attracted to the $ab$ phenotype which can only arise from the $aabb$ genotype because the $A$ or $B$ alleles are always dominant. Nevertheless, some gametes from the $Aabb$ and $aaBb$ genotypes can produce $aabb$ genotypes which will then be attacked by the predator thereby surpressing any reversion of $Aabb$ and $aaBb$ to the $aabb$ genotype.

Further simulations are required to find the lowest predation rate ($D_P$) at which the establishment of new colonies becomes possible.

5.5.3 Varying the general death and predation rates (PR18-21)

These simulations are similar to PR01, above, but investigate the interaction of the general death rate ($D_0$) and the predation rate ($D_P$) on the progress of the Dobzhansky-Müller mechanism.

Fuller details of the input and output data are given in Appendix C.5.3 on page 315.

5.5.3.1 Varying the general death and predation rates (PR18-21): input

The input assumptions are similar to those in PR01 with initial small (2x2) corner colonies of the mutant genotypes $Aabb$ & $aaBb$. The predator is attracted to the $ab$ phenotype which arises only from the genotype $aabb$ and this type constitutes the majority of the initial population. The $AB$ phenotypes are excluded at inception as non-viable (rather than by predation).

The general death rate, which applies to all organisms is variously set at 0%, 0.5%, 1%, 2%, 3%, 4%, 10%, 20%, and 40% while the predation rate is set at intervals of 0.1% between 0.0% and 1.0%.

Earlier, short simulations (not reported here) indicated that the model responds erratically with these settings and therefore 50 reruns were made with each set of assumptions in an attempt to distinguish the general trends in performance.

The simulations were only run for 2000 time-steps because this was found to be sufficient to see if a population of mutants had become established.
Figure 5.39: Mean population of AAbb & aaBB at t=2000 for various general death and predation rates (PR18-21)
5.5.3.2 Varying the general death and predation rates (PR18-21): results

The results from these simulations seem to include a great deal of random variation which made their interpretation difficult. Various analyses of the results were tried in order to extract any trends and the development of the population of the mutant genotypes $AAbb$ & $aaBB$ seems heavily dependent on the early survival of the seed colonies: where the seed colonies are overwhelmed there is little chance of their being reinstated and consequently the outcome is very sensitive to the early development of the mutant colonies.

Various attempts to discern trends were tried, including looking at the standard deviation of the populations in the 50 sets of results. The clearest trends were shown by the mean (of 50 runs) of the total number of individuals with two mutations, i.e. the genotypes $AAbb$ & $aaBB$. These results are expressed as absolute population numbers rather than as percentages because some of the percentages are very small.

The summary results are given in Table 5.18 on page 143 and the mean populations after 2000 time-steps are plotted against the general death rate ($D_0$) in Figure 5.39 on page 142.

These figures show that highest incidences of the $AAbb$ and $aaBB$ genotypes develop when the predation rate ($D_P$) is highest, but this is limited by the prevailing value of the general death-rate ($D_0$).

<table>
<thead>
<tr>
<th>$D_P$:</th>
<th>0.0%</th>
<th>0.1%</th>
<th>0.2%</th>
<th>0.3%</th>
<th>0.4%</th>
<th>0.5%</th>
<th>0.6%</th>
<th>0.7%</th>
<th>0.8%</th>
<th>0.9%</th>
<th>1.0%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0%</td>
<td>0.0</td>
<td>2.3</td>
<td>5.6</td>
<td>13.4</td>
<td>26.3</td>
<td>40.6</td>
<td>57.3</td>
<td>89.1</td>
<td>118.7</td>
<td>138.2</td>
<td>204.1</td>
</tr>
<tr>
<td>0.5%</td>
<td>1.0</td>
<td>7.4</td>
<td>15.0</td>
<td>22.2</td>
<td>43.9</td>
<td>71.6</td>
<td>98.4</td>
<td>149.3</td>
<td>200.8</td>
<td>250.0</td>
<td>317.1</td>
</tr>
<tr>
<td>1.0%</td>
<td>1.5</td>
<td>2.8</td>
<td>7.9</td>
<td>15.7</td>
<td>34.4</td>
<td>67.9</td>
<td>93.0</td>
<td>105.7</td>
<td>151.0</td>
<td>226.3</td>
<td>382.4</td>
</tr>
<tr>
<td>2.0%</td>
<td>0.9</td>
<td>3.7</td>
<td>4.0</td>
<td>17.2</td>
<td>16.4</td>
<td>59.0</td>
<td>86.4</td>
<td>111.2</td>
<td>132.8</td>
<td>224.2</td>
<td>241.4</td>
</tr>
<tr>
<td>3.0%</td>
<td>1.2</td>
<td>0.0</td>
<td>1.7</td>
<td>13.6</td>
<td>11.7</td>
<td>32.8</td>
<td>52.6</td>
<td>46.2</td>
<td>54.1</td>
<td>140.5</td>
<td>176.7</td>
</tr>
<tr>
<td>4.0%</td>
<td>0.0</td>
<td>0.9</td>
<td>0.9</td>
<td>15.3</td>
<td>19.2</td>
<td>37.7</td>
<td>26.1</td>
<td>68.7</td>
<td>97.6</td>
<td>100.9</td>
<td>111.8</td>
</tr>
<tr>
<td>10.0%</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>2.8</td>
<td>0.0</td>
<td>22.3</td>
<td>12.3</td>
<td>48.1</td>
<td>83.6</td>
<td>48.5</td>
<td>36.5</td>
</tr>
<tr>
<td>20.0%</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>14.1</td>
<td>16.3</td>
<td>0.0</td>
<td>71.1</td>
<td>23.9</td>
<td>110.5</td>
<td></td>
</tr>
<tr>
<td>40.0%</td>
<td>2.8</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>32.5</td>
<td>56.4</td>
<td>0.0</td>
<td>33.0</td>
<td>33.0</td>
<td>0.0</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.18: Mean population of AAbb & aaBB at $t=2000$ for various general death ($D_0$) and predation rates ($D_P$) (PR18-21)

5.5.3.3 Varying the general death and predation rates (PR18-21): conclusions

In general higher values of predation are better at facilitating the speciation process and a value of 1% was the highest and most successful value used in these simulations. Where the general death-rate ($D_0$) is less than or equal to the predation rate, the predation seems able to facilitate the speciation process. These results are subject to considerable random error.
and further investigation would be required to determine a more robust relationship between the general death rate and the predation rate. However, when the general death-rate becomes substantially larger than the predation rate ($D_P$) the influence of the predation diminishes and the speciation process is less successful.

It appears that sympatric speciation can be achieved by the Dobzhansky-Müller mechanism once appropriate initial mutations have occurred and there is some external predator to deplete the pre-existing population.
Chapter 6

Conclusions

This chapter draws together the ideas from the previous chapters and is divided into the following four sections:

- In 6.1 the appropriateness of the selected methodologies is reviewed.
- In 6.2 the principal conclusions from the simulations are brought together.
- In 6.3 the overall conclusions about sympatric speciation are summarised.
- In 6.4 possible avenues for future research are discussed.

6.1 Review of the methodology

Three key tasks for this investigation of sympatric speciation were identified early in the project (see Section 1.1.4 on page 18).

These tasks were:

- defining sympatric speciation
- modelling biological genetic processes
- recognising when speciation occurs

The treatment of these objectives are reviewed in the following sub-sections.
6.1.1 Defining sympatric speciation

Speciation proved difficult to define for all organisms but, in Section 2.1.2.1, it was concluded that, for the purposes of this research, speciation would be defined by the Biological Species Concept. This can be summarised as follows: mutual infertility between sexually reproducing organisms is the indicator of speciation. This definition and hence this research is necessarily confined to sexually reproducing organisms.

The question of what is truly “sympatric” was discussed further in Section 5.2.4 on page 95 where it was concluded that, to be truly “sympatric”, speciation must occur in a uniform environment. This definition excludes environments with any variation of conditions within them.

6.1.2 Validity of the modelling

As stated in the introduction (Section 1.1.4 on page 18) the objective of this research is to investigate the logical possibility of sympatric speciation in nature, rather than to model any particular biological examples. This approach was only considered valid if the modelling process realistically represents the behaviour of biological systems.

The *Sympatricia* model, used for this research, was based on a modelling concept devised by Penna [1995]. Derivatives of the Penna model have been used successfully by others: for example, by Stauffer [2007] to represent the incidence of genetic diseases with age and by Cebrat et al. [2012] to investigate gene groupings. These examples indicate that the Penna model can successfully represent real biological processes.

*Sympatricia* simulates a limited range of randomised genetic processes which includes crossovers and individual gene mutations during copying. These features were considered sufficient to address the research objective of investigating sympatric speciation. Other genetic copying errors such as introns and exons were not modelled. Autopolyplody was specifically excluded from the model because it is likely to provide an immediate route to reproductive isolation from the parent population and this project was aimed at investigating more subtle speciation processes.

*Sympatricia* can represent only diploid sexually reproducing organisms which, as explained in Section 2.1.1.1 on page 22, include only a sub-set of all organisms. Nevertheless this sub-set can be considered as a valid group of organisms in which to investigate sympatric speciation.

One area of concern with this type of model is the modelling of multiple genotypes. In Section 4.6 on page 65, it has been shown that the Penna model, in its original form, is not good at modelling multiple species: it tends to revert to a single genotype representing one species. However, after the modifications described in Section 4.6, the *Sympatricia* model was able to retain distinct species for long periods, up to 100000 time-steps.
In addition to the simulations recorded in this thesis, the model was extensively tested to ensure that its mechanisms and algorithms were producing the required effects.

It may therefore be considered that, within the limited scope set for this project, the *Sympatria* model is a reasonably good representation of the genetic interactions of diploid organisms with multiple species present at one time. It therefore provided a useful environment in which to investigate sympatric speciation without autoploidy.

### 6.1.3 Recognising speciation

The recognition that speciation had occurred was approached in three different ways.

Firstly, in modelling the Dobzhansky-Müller mechanism (see Section 5.4 on page 121), an infertility barrier was defined as an input to the simulation. For example, defining a certain gene combination as non-viable effectively established a species boundary between populations with only one or the other of those genes.

Secondly, an unexpected instance of speciation was found (see Section 5.3.8.4 on page 118), called here "phenotype mismatch speciation".

Lastly, in Section 4.5.3 on page 64, potential boundaries between species were identified by measuring the inter-fertility between neighbouring plants and marking where this was relatively low. This technique was successful in highlighting some potential infertility barriers but was not used successfully to identify speciation.

### 6.2 Conclusions from simulations

Having established that the model is adequate for the research objectives, this section summarises the conclusions from the various simulations reported in Chapter 5.

#### 6.2.1 Simulations of the growth in genotype diversity

The first simulations duplicated work by others. They were undertaken at the beginning of the project (see Section 5.1 on page 78) and served to demonstrate that the newly developed version of the model produced results similar to those from Waga et al. [2007]. This provided a reassuring start after which different types of simulation could be attempted.

These simulations demonstrated the "purification" process, which progressively eliminates mutations from the population, as described by Waga et al. [2007]. Later in the research project it was realised that this process indicates that the Penna model is biased in favour of the single unmutated genotype and therefore tends not to allow multiple species to co-exist.
6.2.2 Simulations of environmentally acquired mutations

The first enhancement to the model allowed for the acquisition of genes which provided protection against hostile environments. These simulations, described in Section 5.2 on page 82, show that genes for protective attributes are accumulated. This was not an unexpected result and is well understood as an essential feature of natural selection.

Variations in the environment were introduced in the expectation that some evolutionary bifurcation might arise but no such effects were observed.

More specifically, it is likely these simulations could not have demonstrated sympatric speciation because: (a) there was no separation into reproductively isolated groups and (b) the spatial variation of the environment could be seen as non-sympatric.

6.2.3 Simulations maintaining multiple genotypes

There then followed a series of long-term simulations of up to 100000 time-steps (see Section 5.3 on page 96) to investigate the probability of multiple genotypes coexisting in the model. These confirmed the tendency of the model to favour one genotype, represented by the ‘unmutated’ form.

In order to successfully investigate speciation it was necessary to modify the model to allow more than one genotype to coexist with equal probability of survival. The model was therefore modified by introducing phenotype matching and assortative mating which ensured that similar phenotypes tended to mate. This allowed multiple genotypes to remain stable and separate for many thousands of time-steps.

These simulations also demonstrated the persistence of homozygous genotypes because they always generate gametes which can produce genotypes identical to the parent population. However, the tendency for homozygous genotypes to persist did not preclude the coexistence of distinct homozygous forms.

The model was then ready to investigate possible speciation scenarios without favouring one particular genotype.

6.2.4 Speciation by phenotype mismatching

As mentioned in 6.1.3, above, in the process of investigating multiple coexisting genotypes an unexpected speciation process was observed. Simulations PTAM5 (Section 5.3.7 on page 110) and DLP A0 (Section 5.3.8 on page 115) demonstrate a speciation mechanism named here as “phenotype mismatch speciation”.

Phenotype mismatch speciation was shown to depend on structural incompatibility arising from two alleles, one dominant and one recessive. Whether this phenomenon occurs in nature is
outside the scope of the current research but such a fundamental difference between alleles may be very rare or non-existent. If such a pair of alleles did exist they would necessarily separate into two species and their existence as alternative alleles within one species would no longer be evident.

In Section 5.3.8.4 on page 118 it was demonstrated that “phenotype mismatch speciation” is substantially different from the Dobzhansky-Müller mechanism.

6.2.5 Simulations with the Dobzhansky-Müller mechanism

Having ensured that multiple genotypes could coexist in the model, several simulations of the Dobzhansky-Müller mechanism were attempted using different starting populations and different probabilities of mutation, as described in Section 5.4 on page 121.

The Dobzhansky-Müller mechanism starts from a uniform parent population in which a small number of mutant individuals exist. The population then develops into two incompatible groups (species) through genetic changes at two loci while an intermediate gene combination is deemed to be non-viable. This process is described more fully in Section 2.3.2 on page 31.

In the simulations, the original population was characterised by unmутated genes at two loci, represented as genotype \( aabb \) and the separate species as genotypes \( AAbb \) & \( aaBB \) while any coincidence of the mutations \( A \) & \( B \) is deemed to be non-viable.

In the initial simulations, it was found that small initial colonies of mutants (for example: \( Aabb \) & \( aaBb \) ) frequently died out because they were overwhelmed by the pre-existing \( aabb \) genotypes.

However, by using a non-zero rate of continual mutation, the speciation process succeeded, but the continual insertion of mutations might not be considered a proper implementation of the Dobzhansky-Müller process which should proceed from a single occurrence of the mutants \( Aabb \) & \( aaBb \).

Other methods were therefore tried to facilitate the development of the mutant colonies.

6.2.6 Simulations of Dobzhansky-Müller with predators

As reported in Section 5.5 on page 135, various predators were introduced into the environment where the Dobzhansky-Müller mechanism was modelled.

In the first such simulation, see Section 5.5.1 on page 135, predators were introduced to exclude the \( AB \) phenotypes. In this case, the initial population used was equally divided between the mutant heterozygous genotypes \( Aabb \) and \( aaBb \) and separate populations of genotypes \( AAbb \) and \( aaBB \) developed. However, these simulations represented only part of the Dobzhansky-Müller mechanism because they did not start from the completely unmутated \( aabb \) genotype.
The earlier Dobzhansky-Müller simulations had already demonstrated that small initial mutant colonies in a majority unmutated population generally died out, and therefore a different approach was sought. Instead of introducing a predator to exclude the $AB$ phenotype, a different predator was introduced to deplete the original $aabb$ population in order to allow the small colonies of mutant genotypes ($Aabb$ & $aaBb$) to overcome the pre-existing $aabb$ population and become established. These simulations were successful in modelling the Dobzhansky-Müller process and demonstrated that a predator can facilitate the process after the introduction of small localised colonies of mutants.

Further simulations established that the predation rate generally had to be higher than the general environmental death rate for the effect of the predator to be significant. Within this limit the Dobzhansky-Müller speciation could be demonstrated in a sympatric environment.

It would therefore appear that speciation is facilitated by a predator which depletes the original unmutated population thereby allowing the mutated species to develop in a sympatric environment.

6.3 Overall conclusions

Using a modified form of the Penna model, sympatric speciation was demonstrated using the Dobzhansky-Müller mechanism.

The simulations showed that in order for the Dobzhansky-Müller process to achieve speciation some external agency, possibly a predator or disease, was needed to deplete the existing population so that new species could become established. Simulations of this mechanism do not appear to have been reported elsewhere in the literature, despite the Dobzhansky-Müller mechanism having been first described in the 1930s.

A different mechanism for speciation was also found and named “phenotype mis-match speciation”. This was shown to depend on structural incompatibility arising from two alleles, one dominant and one recessive and was distinct from the Dobzhansky-Müller mechanism.

These two findings seem to suggest that sympatric speciation could arise through genetic processes without autopolyplody but whether these genetic mechanisms actually occur in nature is beyond the scope of the current research.

6.4 Possible future developments

There are a number of areas where future research in this area might be productive.
6.4.1 Predations and mimicry

The interaction of predation, mimicry and hybrid speciation is probably the most fruitful area for further investigation. This could be developed by using a Dobzhansky-Müller type of process in conjunction with the field research into Heliconius butterflies, by Jiggins [2006], Kronforst et al. [2013] and Smith and Kronforst [2013], as described in Section 2.3.3 on page 32.

6.4.2 Spontaneous colony development

The colony boundary identification algorithms developed in this project might be used to identify potential species separations in less structured populations than used in the Dobzhansky-Müller simulations undertaken in the current research.
Appendices
Appendix A

Details of the Sympatria model

This appendix provides additional detail of the input, output and operation of the Sympatria model which is discussed in Chapter 4 of the thesis.

The appendix is divided into three parts:

- A.1: Input to the model;
- A.2: Output mechanisms;
- A.3: Technicalities of the model.

A.1 Input to the model

The Sympatria model is controlled by a text input file which can set the input parameters in any order. Variables which are not set are given default values which are coded within the model. Any input parameter may be given multiple values in the input file so that the model will be run several times varying the input parameter as specified. Where more than one parameter is given multiple values, all combinations of values are used with the parameter given last in the input file changed most frequently. The input file may also set the number of times the model is rerun with each combination of input parameters. The same set of random number seeds is used for each groups of reruns to maintain consistency of approach.

This form of input allows multiple batches of simulations to be run unattended over long periods.

The input data also specifies the output of ‘reports’ which are generally tables of data from which graphs are drawn.

Lines of input data starting with ‘//' are comments which are ignored by the model. Comments can also be included at the end of data line after the final semicolon.
A.1.1 Input parameters

The following lists shows all the available input parameters and their default values, marked ‘df’. Parameters for special effects, such as ageing and assortative mating, have default values which switch off their effect. This allows the relevant section of code to be disregarded during execution thus speeding up processing.
//Sympatria input file
Computer: Toshiba Satellite T110-107 Laptop;

//TITLES
Purpose: TEMPLATE with default inputs; //df = default values
Prefix: TEM_; //df resorts to prefix on filename xx else fails
OutFolder: .; //df .

//LATTICE
Rows: 3; //df 3
Cols: 3; //df 3
WrapRows: No; //df No
WrapCols: No; //df No
SwapsPerCycle: 0.10; //df 0.10

//INITIAL SETUP df 'UniformZeroComp'
InitialSet: UniformZeroComp;

//BASE DATA
MateRange: 1; //df 1 (deprecated form 'Mate')
SeedRange: 1; //df 1 (deprecated form 'Baby')
MatureAge: 1; //df 1
Seeds: 1; //df 1 (deprecated form 'Babies')
Recomb: 0.480; //df 0.48
Crossovers: 1.000; //df 1
Mutation: 0.001; //df 0.001

//AGEING (df FALSE)
BirthCritical: 0; //df 0 (deprecated form 'BirthLoci')
LociPerYr: 0.0; //df 0.0
MutTolerance: 64; //df 64

//PHENOTYPE FATALITY (a fatal combination of mutations)
PhenotypeFatal: 0x0000000000000000; //df ZERO (implies not used)

//PHENOTYPE COMPATIBILITY
PhenotypeLoc: 0x0000000000000000; //df ZERO (implies not used)
PhenoTolerance: 0; //df 0 (64 implies not used)

//ASSORTATIVE MATING
AssortativeLoc: 0x0000000000000000; //df ZERO (implies not used)
AssortativeIndex: 0; //df (min0, max100) increases influence of AM. df 0 (implies not used)

Table A.1: Input data options (sheet 1 of 3)
//PREDICTORS
//Preditor death rate is added to EDR
PredationRate: 0.0 //df 0.0 (implies not used)
PredAttractLoci: 0x0000000000000000; //loci of interest df 0x0000000000000000
PredAttractGenes: 0x0000000000000000; //attractive pattern df 0x0000000000000000
PredDeterLoci: 0x0000000000000000; //loci of interest df 0x0000000000000000
PredDeterGenes: 0x0000000000000000; //pattern which is deterrent df 0x0000000000000000

//ENVIRONMENT
//EnvColBands: 1; //df 1
//EnvRowBands: 1; //df 1

//VARIATION OF ENV VALUE
EnvFunction: [none]; //df [none] sets EnvAction=FALSE
//EnvFunction: ColBands;
//EnvFunction: LinearByCol;
//EnvFunction: LinearByRow;
//EValue1: 0.0; //df 0.0
//EValue2: 0.0; //df 0.0
EVgradient: 0.0; //EVperCol on central EV optimum //df 0.0

//ENVIRONMENTAL DEATH
EnvDeathRate: 0.01; //df 0.01

//PLANT response to Environment
//EVOptimum: 0.0; //df 0.0
//EVIncrLoci: 0x0000000000000000; //df ZERO
//EVDecrLoci: 0x0000000000000000; //df ZERO
//EVIncPerMutation: 0.0; //df 0.0
//EVDecrPerMutation: 0.0; //df 0.0
//EDRperEV: 0.0; //df 0.0

//STEP CHANGE EVENT
//EnvEventAt: 0; //df 0 sets EnvEvent=FALSE
//EnclaveCols: 0; //df 0

//CYCLES
Generations: 20; //df 20
Stops: 20; //excluding the report at time 0 //df 20
//GenStep: 0; //df 0 used for steps between stops instead of Stops
GeometricSteps: No; //df No
ReRuns: 1; //df 1
RandomSeed: 0; //df 0 (deprecated form 'Seed')

//SHOW OPERATIONS
Debug: No; //df No
ShowEnvDeaths: No; //df No
ShowGeneDeaths: No; //df No
ShowSeedGenes: No; //df No
ShowPopS: No; //df No. Used in rptP
ShowCompatibility: No; //df No. Used in rptQ and rptKs

//REPORTING
ReportsZoned: No; //df No

//rptA: Age distribution
ReportAges: No; //df No
AgeGroups: 10x10; //df 10x10 as (No of Groups) x (Range in each)

//rptB: Local Compatibility against Mutation across lattice from model
ReportCompMutation: No; //df No

//rptC: Compatibility across lattice
ReportCompatibility: No; //df No

//rptD: Death by age-group
ReportDeaths: No; //df No

Table A.2: Input data options (sheet 2 of 3)
Table A.3: Input data options (sheet 3 of 3)
A.1.1.1 Initial population options

The shape and genetics of the initial population are set using the ‘InitialSet’ input parameter. The definitions of all the available initial configuration options are listed below:

- **UniformZeroComp**: the whole lattice is set with one chromosome zero paired with its complement.
- **UniformZero**: the whole lattice is set all chromosomes zero.
- **UniformRandomComp**: the whole lattice is set with the same random chromosome paired with its complement.
- **UniformRandomZero**: the whole lattice is set with one chromosome zero paired with its complement.
- **RandomComp**: every cell in the lattice has one random chromosome paired with its complement.
- **RandomZero**: every cell in the lattice has one random chromosome paired with zero.
- **RandomPair**: two adjacent identical cells at the centre of the lattice with one random chromosome paired with its complement.
- **Random4Pair**: two adjacent identical cells at the corners of the lattice with one random chromosome paired with its complement.
- **Zero4Pairs**: two adjacent identical cells at the corners of the lattice with zero chromosomes.
- **ZeroPair**: two adjacent identical cells at the centre of the lattice with zero chromosomes.
- **ZeroPair00**: two adjacent identical cells at the (0,0) corner of the lattice with zero chromosomes.
- **ZeroPairLeft**: two adjacent identical cells at the mid-left of the lattice with zero chromosomes.
- **RandomCompPatches**: uniform patches approximately the size of the pollination range, of one random chromosome with its complement. This provides a random range of plants more sustainable than a totally randomised population in which there would be minimal probability of finding a compatible mate.
- **0x0000000000000000 0x0000000000000000**: uniform across lattice with chromosomes as given in hexadecimal.
*LeftRight* 0000000000000000_0000000000000000_0000000000000000_0000000000000000: uniform in two halves of lattice with chromosomes as given by four hexadecimal numbers.

*ABfilled* 0x0000000000000001 0x0000000000000002: uniform aabb with the loci of a & b as indicated.

*ABhalves* 0x0000000000000001 0x0000000000000002: lattice in halves (left/right or top/bottom) of Aabb & aaBb.

*ABcorners* 0x0000000000000001 0x0000000000000002 5: uniform aabb with corners colonies of Aabb & aaBb, size as given, e.g. 5x5.

The default option for the initial population is *UniformZeroComp* and this is selected if the option specified in the input file is not recognised.
A.2 Output mechanisms

A.2.1 Console output

When running, the model provides a minimum of progress data to the console. This can be increased by setting various ‘show’ or ‘debug’ options in the input file. These options are useful during development and for checking specific details but they generate large volumes of text which substantially slows the progress of the model. The model is normally run with the minimum of console output.

A.2.2 Reporter objects

A set of the ‘reporter’ objects within the model can be invoked to produce data and graphs plotted from that model. The data used to produce the graphs is tabulated in text files and can be independently analysed. The following summary gives the axes and details of the output graphs produced by each ‘reporter’. The output data is collected at snapshot intervals specified in the input. However, the number of data snapshots is limited to 20 in order to avoid excessively complicated graphs.

A.2.2.1 Age Profile (A)

x-axis: Time
y-axis: Population of live plants, in stacked histogram in age groups. Age grouping bands are set by a parameter.

A.2.2.2 Local Compatibility v Mutation (B)

x-axis: Number of single mutation in chromosomes
y-axis (right): The proportion of adult plants which have each number of single mutations (but not necessarily mutations at the same loci) plotted as a histogram.
y-axis (left): Local Compatibility

The ‘local compatibility’ of a plant is defined as the compatibility which it has with its neighbours within pollination range. This is calculated by pairing a plant with all its potential pollination partners, creating all possible zygotes between them and counting the proportion which would be viable. The resulting proportion is defined as the local compatibility and represents the probability that the parent plant can viably cross with its neighbours. This is plotted against the number of single mutations which the plant has and shown as the average local compatibility for all the plants which have the same number of single mutations.
A.2.2.3 Local Compatibility v Column (C)

x-axis: Columns across the lattice
y-axis: Average local compatibility in each column (see definition above).

A.2.2.4 Death by Age-group (D)

x-axis: Time
y-axis: Number of deaths from all causes by age-group in stacked histogram.
The number of deaths is measured from those plants marked as dead at the time when the sample is taken (at regular time intervals). Each dead plant only exists for one time-step after which it is cleared and the lattice position becomes vacant. Age is measured in time-steps and the size of the age-groups to be displayed is an input parameter. Plants with an age higher than the highest age-group are excluded from the count.

A.2.2.5 Families v Time (F)

x-axis: Time
y-axis: Number of distinct genotypes (Families).
Plants are considered to be in the same family if their genotypes are identical. The typing algorithm is unaffected by the order in which the haplotypes of a plant are stored, thus: if the genetic type of a plant with haplotypes $H_p$ and $H_q$ is given by a function $T(H_p, H_q)$, then $T(H_p, H_q) = T(H_q, H_p)$.

A.2.2.6 Compatibility v Separation (K)

x-axis: Separation of plants
This is the Euclidian distance between each sampled pair of plants.
y-axis: Compatibility of random pairs
Compatibility is measured between a random sample of pairs of plants from across the whole lattice. The number of sample pairs is given as an input parameter: e.g. 100. Compatibility is defined as the proportion of all possible zygotes between the sampled pair which would be viable.

A.2.2.7 Mutations v Locus(M)

x-axis: Loci along the chromosome
y-axis: Proportion of plants which have a mutation at that locus.
The data may be collected for single, double, or all mutations. The mutation data is captured at each time snapshot and shows the development of mutation patterns over time. The same data is plotted as Mutation v Time, see below.

A.2.2.8 Mutations v Time (N)

x-axis: Time
y-axis: Proportion of plants which have a mutation at each locus, with a line plotted for a selected set of loci.
The data may be collected for single mutations, double mutations or all mutations. This is an alternative view of the Mutation v Locus, above.

A.2.2.9 Population (P)

x-axis: Time
y-axis: Total population numbers.
This is a stacked histogram showing the numbers by status of every lattice position: empty, dead, child or adult.

A.2.2.10 Quarter-column Compatibility (Q)

x-axis: Time
y-axis: Relative compatibility of specific columns of plants at 1/4 and 3/4 points across the lattice.
The relative compatibility is measured as the number of viable zygotes as a proportion of all possible zygotes between all the plants in one column with all plants in the other. This is used to measure if the two halves of the lattice are becoming incompatible.

A.2.3 Maps of the lattice

There are also outputs of images of the lattice from the model. These are images of the whole lattice with each plant position represented by a coloured square. Images are built every snapshot time and are not limited in number. The time and run reference are marked on each image. The genotype, compatibility and colony maps are shown on one image.

After each run the multiple images can be combined into an .mp4 moving image file. The movie frame dwell-time is an input parameter. When multiple reruns are used, each rerun may be captured as a separate moving image.
A.2.3.1 Genetic image

The mapped colours are created from the mutations present in either of the two chromosomes. The mapping of loci to colours is defined by three 64-bit masks, which are input parameters, and transfer selected portions of one of the chromosomes to the three components of an RGB colour. As explained in Section 4.3.2.1 a 24-bit colour palette is not sufficient to accurately display all types of 64-bit chromosome. Further details are given in Appendix A.3.1.7.

A.2.3.2 Compatibility image

In addition to the genotype map it is also possible to plot the ‘local compatibility’ of each plant as a colour image. The local compatibility, as defined in Section 4.5.2, for each plant is displayed on a grey-scale ranging from white for 100% compatibility to black for total incompatibility.

A.2.3.3 Colony image

Compatibility estimation can be used to generate colony numbering, as described in Section 4.5.3.1. Where this has been done the colony numbers may be plotted on a lattice map using arbitrary colours for each colony. If the numbering algorithm has succeeded in maintaining the colony numbers over time, the colour maps of colonies will remain stable showing the development of each colony.
A.3 Technicalities of the model

In this appendix some of the computational techniques used in constructing the model are discussed.

The model was coded in C++ under Linux using the gcc compiler version: g++ (Ubuntu/Linaro 4.6.3-1ubuntu5) 4.6.3.

A.3.1 Object structure

The principal objects are shown below in their hierarchy:

- control - controls the time cycles and outputs
  - inputset - reads and decodes the input data file
  - basedata - holds universal data applicable to all ‘cells’
  - model - holds the lattice of ‘cells’
    - cell - each plant as a ‘cell’
  - reporterA - reads the model data and compiles output
  - ... other reporters

Key features of these objects are given in the following sub-sections.

A.3.1.1 control object

The control object is created by the main program and proceeds to execute the whole modelling process. The control first calls inputset which reads the input file and then repeats model runs according to the input requirements.

For each model run control refers to inputset and adjusts the modelling parameters as required, then runs the model in stages stopping to call reporter objects at the specified intervals.

A.3.1.2 inputset object

When this object is created it reads the specified input text file and decodes each line. The inputset simply records each parameter name and the text of the one or more values assigned to it.

The inputset cycles through all combinations of the input values given, those entered later in the input file changing faster. Thereby, during any model run inputset can provide the current
values of all parameters on request. The next set of values is set when control requests more data.

During the running of the model other objects refer to inputset for current values of parameters, calling them by their text names as shown in the input file. This is an infrequent process and so using text names does not create much delay but using text names simplifies the preparation of the input files.

The input parameters which are varying during a set of model runs are listed by inputset so that they can be reported in various output reports in order to clarify the changes being made.

A.3.1.3 model object

The model contains the lattice as a two-dimensional array of cell objects which is allocated memory dynamically to suit the lattice dimensions specified in the input.

The model also controls the functions which extend across the lattice, for example: setting up the initial plant population and colony recognition and numbering.

A.3.1.4 basedata object

This holds data which is universal across the lattice such as environmental parameters, recombination probability etc. It extracts this data from the inputset object and is visible to all ‘cells’ and the ‘model’.

A.3.1.5 cell object

Strictly speaking the cell object is not a biological entity but a location in the lattice which may or may not be occupied by a plant. The status of each cell is indicated by a user-defined variable ‘STATE’ which may be: empty, blocked, dead, immature, or mature. Only the latter three of these states indicated the presence of a plant.

The cell holds the genetic information about any plant which it may contain. The cell contains methods including functions to generate gametes and mate with other cells.

A.3.1.6 reporter objects

There are a number of different reporter objects; they are listed in Appendix A.2.2.

Each reporter interacts with the the input set to finds if it is required and to obtain its operating parameters. The reporters are individually accessed by the control object but in hindsight it would have been better to create a single reporter with subordinate objects to execute different output.
A.3.1.7 Colour mapping of genotypes

Any 8 loci in the chromosome can be used to drive each colour component. The numerically larger chromosome is always used for the mapping. This has the advantage of ensuring that genotypes with equivalent chromosomes would appear as the same colour, so AB and BA might both have their colour defined by A.

This technique was adequate when dealing with the Penna model in its original form where double mutations at one locus were unusual, but in later versions of the model it became interesting to distinguish single and double mutations. To achieve this the user input allow various options for dealing with the two chromosomes. These are: HI, LO, AND, OR, and SPLIT. The first two simply take the numerically larger or smaller chromosome for the mapping. The next two logically combine the two chromosomes, but these options are of limited usefulness. The last option splits the colour component between the two chromosomes so that single and double mutations appear as different colours. This was achieved by assigning bit 0, 2, 4 & 6 of the colour number to one chromosome and bits 1, 3, 5 & 7 to the other.

A.3.2 Performance of the model

The programming techniques used were principally aimed at enhancing the speed of operation of the model. The lattice can contain thousands of plants. During each time step various operations are performed on all the plants in the lattice and this potentially could take a long time. Consequently, care was taken in writing the code to ensure that it would run as quickly as possible. Some of the techniques used are outlined below.

A.3.2.1 Computing environment

In order to avoid the operational ‘overhead’ of some programming languages, C++ was selected because it allows the closest accessible coding language top assembler code. This was run under Linux initially version 10.04 and subsequently version 12.04. Linux was selected because it offers a wide range of software at no cost to the user. The compiler used was g++ (Ubuntu/Linaro 4.6.3-1ubuntu5).

Graphs were produced by gnuplot, available free under Linux. The input data and plot specification files for gnuplot were generated by C++ code within the model.

Some post-processing of data was done in spreadsheets using LibreOffice 3.5.7.2.

A.3.2.2 Programming in C++

The speed of execution was enhanced by using C++, but one disadvantages of this low-level language is that routines have to be included to execute standard tasks such as reading and
writing data, and encoding and decoding numeric strings. A set of functions was developed for these types of operation.

A.3.3 The user interface

In its first version Sympatria was built with a full graphical user interface (using the Code::Blocks development environment) which allowed interactive alteration of parameter values and a ‘live’ view of the gene-map while the model was running showing a coloured representation of the genes of every cell on the lattice. However, this interface was burdensome to maintain and tended to obstruct the development of additional, useful functionality and it did not provide sufficient numerical output for serious analysis. After discussions with Waga in Wroclaw in October 2011, the model was rebuilt as Sympatria2 reusing some of the original class objects but employing text-based input and output methods. In particular an input text file was introduced to allow multiple batches of runs to be made with ranges of parameter values with controlled sets of random numbers and multiple reruns to allow averaged results. This allowed the model to run numerous alternative examples unattended, typically for several hours overnight. The input files also provided a permanent record of input parameters for future reference. In this version ‘reporter’ objects have been developed to provide detailed numerical output for further analysis. The visual image of the gene-map was still available through the ‘movie’ object which created the map images and combined them into a .mp4 video file which was still useful for reviewing the general performance of the population during any run.

A.3.4 Lattice structure and random access

The principle object in the Sympatria system is the ‘model’, as described above. This contains and controls the lattice of ‘cell’ objects and operates the interactions between the plants which may occupy the cells of the lattice.

A.3.4.1 Dynamic memory allocation

All the major blocks of memory are allocated dynamically and this applies especially to the lattice of ‘cell’ objects. This allows complete freedom in the shape and size of the lattice. The model holds a two-dimensional array of cell objects which is reallocated whenever the lattice size is changed. In parallel with this 2D array there is a one-dimensional vector of pointers to the lattice cells. This is used for two reasons: firstly it offers a quick method of addressing the cells avoiding the additional time taken to calculate the address in two dimensions; secondly, it allows the cells of the lattice to be addressed in a random order.
A.3.4.2 Random access to cells in the lattice

Early in the project it became clear, by observing changes in the map of genes over many generations, the there was a bias in the development of successive generations. As time progressed it appeared that successive generations of similar genotypes would migrate across the lattice in the direction of increasing array indices. This effect is thought to have been a result of the methodical scanning of the lattice by columns and rows indices which could cause the preferential placement of offspring in the direction of greater index because a vacant lattice position is more likely to be found there, other neighbouring positions having already been filled by previously scanned parent cells.

To solve this bias a system of randomised processing to the cells was devised. This method must access every cell in the lattice once during every time-step but needs to do this in a random sequence. A vector of pointers to the cells is therefore set up when the lattice is formed, but is then randomised by swapping random pairs of pointers in the vector. During operation of the model a proportion (specified at input but usually 10%) of the population of cells are swapped at random at every time step. Using these methods the migration of genotypes across the lattice was no longer observed.

A.3.5 Gaining speed by using of memory

The relatively large memory available in modern computers made it possible to employ memory in some processes to enhance the speed of the algorithms.

A.3.5.1 Lists of neighbours

A frequently repeated operation in the operation of the model is searching and identifying suitable neighbouring cells as pollination partners or locations for offspring. The neighbours are defined as those within a given range of the parent cell and the range may be different for pollination and placement. The neighbours of each cell remains constant for a given lattice shape and size and are therefore constant throughout a model run. Every cell object therefore holds pointers to all its potential pollinators and seed positions which can then be quickly accessed (by a random selector) without having to compute the relative positions of neighbours in the lattice array.

This is an example of the use of memory, for the list of neighbours, being used to enhance the speed of an algorithm.
A.3.5.2 Viability of possible gametes

Another calculated value which can be stored is the local compatibility between a plant and its potential mates and it is only necessary to revise this when a potential mate dies or is replaced. This is managed by setting a flag to indicate whether the stored value of the ‘local compatibility’ is current or not. When any plant dies this flag is switched off for itself and all its potential mates.

When local compatibility is called for and is not current it is recalculated and the 128 possible gametes from the given plant are compared with the 128 possible gametes from each adult plant in the pollination range. With each pollination partner there are $128^2$ gamete-to-gamete combinations and, for example, with 8 possible partners within a pollination range of 1, there are therefore 131,072 ($2^{17}$) possible resulting zygotes.

Each plant (cell) can then retain the value of its local compatibility and recalculate it relatively infrequently. The normalised compatibility figures do not persist in the same way and must be recalculated every time-step as the maximum and minimum compatibilities across the lattice are expected to change continually.

In Section 4.5.2, it is shown that the probability of producing a gamete without a crossover is $(1 - P_R)/2$ and of producing one by crossover is $P_R/126$, where $P_R$ is the probability of crossover. These probabilities remain constant throughout each model run, because $P_R$ is a universal constant, and could therefore be calculated once and stored for reference in the localcompatibility algorithm.

A.3.5.3 Randomised probability function

Another speed-enhancement was used to achieve a random response with a predefined probability. For example, the probability of a crossover occurring ($P_R$) was input as a fraction but immediately converted to an integer $I_R = P_R \cdot R_{max}$, where $R_{max}$ is the number of random integers produced by the random number generator. Subsequently, the decision whether or not to include a crossover was simply decided by comparing the next random number with the stored integer $I_R$. Similar mechanisms were used where possible for all randomly driven decisions.
Appendix B

Theoretical analysis

B.1 Probability of viability in the Penna model

This appendix summarises some theoretical analysis of the principles underlying some of the genetic processes being modelled. The symbols used in this section are summarised at the beginning of this thesis.

B.1.1 Probability of viable offspring

A problem experienced with the current model is the low probability of compatibility between identical plants and the consequent lack of cohesion of colonies of similar plants. Consider the case where a number of mutations have accumulated in both haplotypes, and the recombination rate is non-zero (i.e. no crossovers) and two identical parent plants have the haplotypes \([H_p, H_q]\). The possible zygotes from these parents would be \([H_p, H_p]\), \([H_p, H_q]\), \([H_q, H_p]\), and \([H_q, H_q]\), with only two of these \([H_p, H_q]\) and \([H_q, H_p]\) being viable, the other two having coincident mutations in their identical haplotypes. The probability of viable offspring in these circumstances is therefore only 0.50. This seems unrealistic and only a marginal increase would seem to result from introducing a crossover.

B.1.2 Probability of successful crossover

During reproduction, let us assume that each parent provides a gamete derived from its own two haplotypes by a random crossover process without the random addition of any new mutations. If mutations appear at the same locus in the gametes from both parents then the offspring is, by definition, non-viable. This raises the question: what is the probability that two given parents will produce a viable offspring? The following analysis demonstrates that, in the case of one crossover, the probability of producing a viable offspring is \(0.75^{M_z}\), where \(M_z\) is the number of mutation bits common to the gametes from the parents.
Assume that parents A and B have haplotypes: $H_{A0}$, $H_{A1}$, $H_{B0}$ and $H_{B1}$ respectively. For example using 32-bits as an illustration:

$$
H_{A0} = 00000101010100101010101010101000 \\
H_{A1} = 01110010001001000000000001010011
$$

(B.1)

The numbers of mutations (bits) in each haplotype of parent ‘A’ are:

$$
M_{A0} = \beta(H_{A0}) \\
M_{A1} = \beta(H_{A1})
$$

where the function $\beta(h)$ counts the number of bits in bit-string $h$.

The bits never coincide at a locus in one parent because if they did the parent cell would not have been viable, therefore the total numbers of mutations in parents A and B are:

$$
M_A = M_{A0} + M_{A1} = \beta(H_{A0} \cup H_{A1}) \\
M_B = M_{B0} + M_{B1} = \beta(H_{B0} \cup H_{B1})
$$

A gamete ($H_{A0}$ x $H_{A1}$) produced from a single crossover in parent ‘A’, will in general have $M_{GA}$ mutations where $M_{GA} \leq M_A$ because some bits may be omitted in the crossover but no more can be generated. Similarly, the gamete from B will contain $M_{GB}$ mutations where $M_{GB} \leq M_B$.

Putting these gametes together to form a zygote ($Z$) in which $M_z$ mutations may coincide and:

$$
M_z \leq M_{GA} \\
M_z \leq M_{GB}
$$

Only at these $M_z$ loci can fatal coincidence of mutations occur and at each of these loci 4 outcomes of the crossover processes are possible: 00, 01, 10 and 11. Of these only one combination (11) is fatal. It can be shown that a random crossover algorithm in the model makes each of these four outcomes equally probable and therefore, considering these $M_z$ loci, the probability of a viable outcome is $P_v$, given by:

$$
P_v = \frac{3^{M_z}}{4^{M_z}} \\
P_v = 0.75^{M_z}
$$

There is therefore an exponential decrease in probability of viable outcome as the number of coincident bits increases, as shown in the Table B.1:
It is also interesting to note that because the number of coincident mutations cannot be more than the number of mutations in the gametes, it follows that:

\[ M_z \leq M_{GA} \Rightarrow P_v \geq 0.75^{M_{GA}} \]
\[ M_z \leq M_{GB} \Rightarrow P_v \geq 0.75^{M_{GB}} \]  \hspace{1cm} (B.2)

This is because it is assumed above that the number of coincident mutation bits will always be less than the number of mutations in either parent. However in the model it is possible to find more mutations in a gamete than in its parent by the addition of copying error mutations during meiosis.

Taking in conjunction with the theory above, the probability of a viable outcome, between identical parents, is only slightly increased above 0.50 by the introduction of a crossover.

### B.1.3 Compatibility with a single crossover

In a separate exercise using the Sympatria model with one crossover allowed, it appeared that the the compatibility of similar plants was approximately 0.262. In this context the term compatibility is used to mean the probability that two given parent will produce a viable offspring which is assessed by comparing all possible gametes which the parent might produce and counting the proportion of those pairs which would produce a viable zygote.

The following analysis shows the origin of the figure 0.262 for the compatibility of a pair of identical parents with simple complementary haplotypes of \( N \) genes, for example:

\[ i f \ N = 32 \]
\[ H_0 = 11111111111111111111111111111111 \]
\[ H_1 = 00000000000000000000000000000000 \]

After one crossover, a parent ‘A’ of this type could produce, for example, gametes with 7 or 25 mutations, thus:

\[ G_{A7} = 11111111000000000000000000000000 \]
\[ G'_{A7} = 00000001111111111111111111111111 \]

where \( G' \) indicates the complement of \( G \).
Each parent can produce \( N \) pairs of possible gametes with the crossover being after the 1\(^{st} \) to the \( N^{th} \) locus\(^1\). Introducing the convention that \((p : q)\) indicates a gamete with \( p \) ones followed by \( q \) zeros, the possible gametes and their complements from two parents ‘A’ and ‘B’, with crossovers after genes \( i \) and \( j \) respectively, are defined as follows:

\[
\begin{align*}
G_{Ai} &= (i : 0) \\
G'_{Ai} &= (0 : (N - i)) \\
G_{Bj} &= (j : 0) \\
G'_{Bj} &= (0 : (N - j))
\end{align*}
\]

Considering the four parings of these sets of \( N \) possible gametes, each pairing produces \( N^2 \) possible zygotes represented in the following matrices with each element \((i, j)\) indicating a viable (1) or non-viable (0) outcome for the combination \( i, j \). The total number of viable outcomes of each set of pairs is then given as \( K_{pq} \):

\[
G_{Ai} \times G_{Bj} = (i : 0) \times (j : 0) \rightarrow \begin{bmatrix} 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 0 & 0 \end{bmatrix} \rightarrow K_{AB} = 0
\]

\[
G'_{Ai} \times G_{Bj} = (0 : (N - i)) \times (j : 0) \rightarrow \begin{bmatrix} 1 & 0 & \ldots & 0 & 0 \\ 1 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & \ldots & 1 & 0 \\ 1 & 1 & \ldots & 1 & 1 \end{bmatrix} \rightarrow K'_{AB} = \sum_{i=1}^{N} i
\]

\[
G_{Ai} \times G'_{Bj} = (i : 0) \times (0 : (N - j)) \rightarrow \begin{bmatrix} 1 & 1 & \ldots & 1 & 1 \\ 0 & 1 & \ldots & 1 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 1 \\ 0 & 0 & \ldots & 0 & 1 \end{bmatrix} \rightarrow K_{AB'} = \sum_{j=1}^{N} j
\]

\[
G'_{Ai} \times G'_{Bj} = (0 : (N - i)) \times (0 : (N - j)) \rightarrow \begin{bmatrix} 0 & 0 & \ldots & 0 & 1 \\ 0 & 0 & \ldots & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & 1 \\ 1 & 1 & \ldots & 1 & 1 \end{bmatrix} \rightarrow K'_{A'B'} = N + N - 1
\]

\(^1\)A gamete with crossover after the \( N^{th} \) locus is equivalent to no crossover having occurred but is a valid possibility in this analysis.
The total number of viable outcomes $K$ is therefore:

$$K = K_{AB} + K_{A'B} + K_{A'B'} + K_{A'B}$$

$$= 0 + 2\sum_{i=1}^{N} i + 2N - 1$$

$$= 2N(N+1)/2 + 2N - 1$$

$$= N^2 + 3N - 1$$

The total number of possible outcomes (viable and non-viable) is $4N^2$, therefore the probability $P_v$ of a viable outcome is given by:

$$P_v = \frac{K}{4N^2}$$

$$= \frac{N^2 + 3N - 1}{4N^2}$$

$$= 1/4 + 3/(4N) - 1/(4N^2)$$

For $N = 64$, as in the ‘Sympatria’ model, $P_v = 0.261657715$ which corresponds with the value of 0.262 observed. The variation of $N$ with $P_v$ is shown in Figure B.1 and it can be seen that with more than 200 genes the probability of viability tends closely to 0.25.

Figure B.1: Probability of viable offspring v Genes on Haplotype
B.1.4 Gamete compatibility & Hamming distance

It is suggested by Waga et al. [2007] that, in the Penna model, the Hamming distance between the binary numbers holding parental genotypes is a measure of species separation.

Consider a Penna model where compatibility of the gametes depends on there being no coincident mutations at \( L_c \) critical loci of gametes A & B which contain \( m_A \) & \( m_B \) mutations respectively. In order to be compatible \( m_A + m_B \leq L_c \) and none of the mutations must coincide. The Hamming distance between the gametes if therefore given by \( H_{AB} = m_A + m_B \).

The number of configurations of the mutations in A is given by:

\[
N_A = \frac{L_c!}{(L_c - m_A)!m_A!}
\]

In order to maintain compatibility the mutations in B may only occur in the \((L_c - m_A)\) loci which are not opposite mutations in A. Therefore the number of possible configurations of the mutations in B is given by:

\[
N_B = \frac{(L_c - m_A)!}{(L_c - m_A - m_B)!m_B!} = \frac{(L_c - m_A)!}{(L_c - H_{AB})!m_B!}
\]

Consider now changing the Hamming distance \( H_{AB} \) while maintaining compatibility between A&B. This can only be achieved by changing the genes at any of the loci in B which are not opposite mutations in A. This means any of \((L_c - m_A)\) loci may be changed.

Adding mutations to the unmutated loci in B would increase \( H_{AB} \) by adding most differences, but removing a mutation in B would reduce \( H_{AB} \) by 1. There are \( m_B \) mutations in B so the fraction of \( N_B \) which can be changed to give a compatible gamete with \( H_{AB} \) reduced by 1 is \( m_B/(L_c - m_A) \).

The number of gametes created by this change is given by:

\[
N_B' = \frac{m_B}{(L_c - m_A)} \cdot N_B = \frac{m_B}{(L_c - m_A)} \cdot \frac{(L_c - m_A)!}{(L_c - m_A - m_B)!m_B!} = \frac{m_B}{(L_c - m_A - m_B)!m_B!} \cdot \frac{(L_c - m_A)!}{(L_c - m_A - m_B)!m_B!} = \frac{m_B}{(L_c - m_A - m_B)!m_B!} \cdot \frac{(L_c - m_A)!}{(L_c - m_A - m_B)!m_B!} = \frac{(L_c - m_A - m_B)!m_B!}{(L_c - m_A - 1 - (m_B - 1))(m_B - 1)!}
\]

Considering all configurations of the mutations in A:

\[
\sum_{m_A} N_A = \sum_{m_A} \frac{L_c!}{(L_c - m_A)!m_A!} = 2^{L_c}
\]
Similarly:

\[
\sum_{m_B} N_B = \sum_{m_B} \frac{(L_c - m_A)!}{(L_c - m_A - m_B)!m_B!} = 2^{L_c - m_A}
\]

and

\[
\sum_{m_B} N'_B = \sum_{m_B} \frac{(L_c - m_A - 1)!}{(L_c - m_A - 1 - (m_B - 1))!(m_B - 1)!} = 2^{L_c - m_A - 1}
\]

The number of compatible gametes \(B\) has halved from \(2^{L_c - m_A}\) to \(2^{L_c - m_A - 1}\) when the Hamming distance between \(A\) and \(B\) has been reduced by 1.

In the case of an exactly complementary pair, the Hamming distance \(H_{AB} = L_c\), and this compatible pair is the only pair with this separation. With every unit decrease in \(H_{AB}\) the probability of finding a compatible gamete is halved, starting from 100% when \(H_{AB} = L_c\).

This illustrates that as Hamming distance increases the probability of compatibility increases thereby showing that a small Hamming distance is a poor indicator of species similarity when species similarity is measured by gamete compatibility.

**B.1.4.1 Numerical analysis of Hamming distance and compatibility of gametes**

In order to verify the above analysis, all possible pairs of 16-bit gametes were compared and the proportion of compatible pairs was evaluated for each increment of Hamming separation.

The results of this evaluation are shown in Table B.2 on page 177. This showed that at the maximum Hamming separation (16) all pairs were compatible: the pairs were all complementary and therefore had no coincident mutations. For each unit of decrease in Hamming separation the proportion of compatible pairs reduces by half. At a Hamming separation of zero the proportion of compatible pairs was \(2^{-16}\) because all the pairs are identical matches and only one pair (both zero) are compatible out of \(2^{16}\) possible pairs.
<table>
<thead>
<tr>
<th>Hamming $H_{ab}$</th>
<th>All pairs $X_t$</th>
<th>Comp. pairs $X_v$</th>
<th>Proportion $P_v$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>65536</td>
<td>1</td>
<td>0.00002</td>
</tr>
<tr>
<td>1</td>
<td>1048576</td>
<td>32</td>
<td>0.00003</td>
</tr>
<tr>
<td>2</td>
<td>7864320</td>
<td>480</td>
<td>0.00006</td>
</tr>
<tr>
<td>3</td>
<td>36700160</td>
<td>4480</td>
<td>0.00012</td>
</tr>
<tr>
<td>4</td>
<td>119275520</td>
<td>29120</td>
<td>0.00024</td>
</tr>
<tr>
<td>5</td>
<td>286261248</td>
<td>139776</td>
<td>0.00049</td>
</tr>
<tr>
<td>6</td>
<td>524812288</td>
<td>512512</td>
<td>0.00098</td>
</tr>
<tr>
<td>7</td>
<td>749731840</td>
<td>1464320</td>
<td>0.00195</td>
</tr>
<tr>
<td>8</td>
<td>843448320</td>
<td>3294720</td>
<td>0.00391</td>
</tr>
<tr>
<td>9</td>
<td>749731840</td>
<td>5857280</td>
<td>0.00781</td>
</tr>
<tr>
<td>10</td>
<td>524812288</td>
<td>8200192</td>
<td>0.01562</td>
</tr>
<tr>
<td>11</td>
<td>286261248</td>
<td>8945664</td>
<td>0.03125</td>
</tr>
<tr>
<td>12</td>
<td>119275520</td>
<td>7454720</td>
<td>0.06250</td>
</tr>
<tr>
<td>13</td>
<td>36700160</td>
<td>4587520</td>
<td>0.12500</td>
</tr>
<tr>
<td>14</td>
<td>7864320</td>
<td>1966080</td>
<td>0.25000</td>
</tr>
<tr>
<td>15</td>
<td>1048576</td>
<td>524288</td>
<td>0.50000</td>
</tr>
<tr>
<td>16</td>
<td>65536</td>
<td>65536</td>
<td>1.00000</td>
</tr>
</tbody>
</table>

Table B.2: Probability of compatibility against Hamming separation
Appendix C

Simulation data

These appendices contain the detailed input and output data for the simulations which are referred to in the main body of the thesis.

C.1 Simulations on the growth of genotype diversity

These simulations are described and discussed in Section 5.1 on page 78

C.1.1 Growth of genotype diversity: Input

During these simulations, a cylindrical lattice was used with the circumference of the wrapped cylinder made at least four-times the maximum range at which pollination and seeding will take place. This range was chosen to avoid a plant interfering with the placement of its own seedlings or selection of its own pollinators.

The input data file for this simulation is shown in Table C.1 on page 179.
Table C.1: Growth of genotype diversity: input data
C.1.2 Growth of genotype diversity: Results

The results from this simulation are summarised in Table C.2 on page 180 which shows the numbers of genotypes present at intervals of 2500 time-steps, the actual results having been recorded at time intervals of 500 time-steps. A graph of these results and comments on them are given in Section 5.1 on page 78

<table>
<thead>
<tr>
<th>$P_R$</th>
<th>0.1</th>
<th>0.2</th>
<th>0.4</th>
<th>0.8</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2500</td>
<td>7</td>
<td>14</td>
<td>58</td>
<td>803</td>
<td>948</td>
</tr>
<tr>
<td>5000</td>
<td>10</td>
<td>19</td>
<td>92</td>
<td>847</td>
<td>916</td>
</tr>
<tr>
<td>7500</td>
<td>11</td>
<td>25</td>
<td>125</td>
<td>814</td>
<td>754</td>
</tr>
<tr>
<td>10000</td>
<td>11</td>
<td>31</td>
<td>162</td>
<td>750</td>
<td>460</td>
</tr>
<tr>
<td>12500</td>
<td>14</td>
<td>35</td>
<td>191</td>
<td>643</td>
<td>197</td>
</tr>
<tr>
<td>15000</td>
<td>16</td>
<td>41</td>
<td>218</td>
<td>469</td>
<td>84</td>
</tr>
<tr>
<td>17500</td>
<td>18</td>
<td>50</td>
<td>230</td>
<td>291</td>
<td>42</td>
</tr>
<tr>
<td>20000</td>
<td>18</td>
<td>54</td>
<td>231</td>
<td>136</td>
<td>14</td>
</tr>
<tr>
<td>22500</td>
<td>20</td>
<td>57</td>
<td>230</td>
<td>70</td>
<td>5</td>
</tr>
<tr>
<td>25000</td>
<td>23</td>
<td>61</td>
<td>235</td>
<td>48</td>
<td>2</td>
</tr>
<tr>
<td>27500</td>
<td>22</td>
<td>69</td>
<td>243</td>
<td>24</td>
<td>1</td>
</tr>
<tr>
<td>30000</td>
<td>24</td>
<td>73</td>
<td>250</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>32500</td>
<td>25</td>
<td>76</td>
<td>250</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>35000</td>
<td>30</td>
<td>77</td>
<td>245</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>37500</td>
<td>31</td>
<td>81</td>
<td>255</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>40000</td>
<td>32</td>
<td>84</td>
<td>245</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>42500</td>
<td>33</td>
<td>85</td>
<td>242</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>45000</td>
<td>33</td>
<td>89</td>
<td>230</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>47500</td>
<td>36</td>
<td>92</td>
<td>224</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>50000</td>
<td>36</td>
<td>95</td>
<td>223</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>52500</td>
<td>36</td>
<td>105</td>
<td>215</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>55000</td>
<td>39</td>
<td>100</td>
<td>203</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>57500</td>
<td>43</td>
<td>104</td>
<td>203</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>60000</td>
<td>45</td>
<td>107</td>
<td>188</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table C.2: Nos of genotypes with various probabilities of crossover ($P_R$)
C.2 Simulations on environmentally acquired mutations

C.2.1 Simulations in a linear environment

These simulations are discussed in Section 5.2.2 on page 83.

C.2.1.1 Simulations in a linear environment: input

Both cases were generated from the input file ‘EVLN’ shown in Table C.3 on page 182.

The cases discussed were taken from the first and last values, 0.005 and 0.200, used for $\Delta D_V$ (shown as input parameter ‘EDRperEV’ in the input file).
Table C.3: EVLN input data
C.2.2 Simulations in a changing environment

This simulation is discussed in Section 5.2.3 on page 91. It assesses the long-term quarter-column compatibility with the environmental gradient (\( \Delta V_c \)) set to zero.

C.2.2.1 Simulations in a changing environment: input

The input for this simulation ‘EVVR’ is shown in Table C.4 on page 183.

```plaintext
//TITLES
Computer: Laptop;
Purpose: Quarter-column compatibility with Long Period;
Prefix: QCz50k_;
//LATTICE
Rows: 32;
Cols: 128;
WrapCols: Yes;
SwapsPerCycle: 0.25;

//INITIAL SETUP
InitialSet: UniformZero;

//BASE DATA
Mate: 2;
Baby: 2;
MatureAge: 1;
Births: 1;
BirthLocis: 64;
LociPerFrr: 0.0;
MutTolerance: 0;
Recomb: 0.5;
CrossOvers: 1.0;
Mutation: 2;

//ENVIRONMENT
EnvFunction: LinearByCol;
EVgradient: 0.0;
EnvDeathRate: 0.005;
EDRperEV: 0.01;

//PLANT RESPONSE to ENV
EVOptimum: 32.0;
EVincrLocis: 0x0FF0000000000000;
EVincrLocis: 0x0FF0000000000000;
EVDecrLocis: 0x0000000000000FF0;
EVDecrLocis: 0x0000000000000FF0;
EVDecrPerMutation: 2.0;
EVincrPerMutation: 2.0;

//CYCLES
Generations: 50000;
Stops: 20;
GeometricSteps: No;
Seed: 99;
ReRuns: 10;

//REPORTING
ReportQuarterCompatability: Yes;

//MOVIES
GeneImages: Yes;
ScaleMax: 0.1;
ImageDwell: 0.5; //seconds per frame
CaptureReruns: 1; //movies of reruns
//Colour mapping
RedMask: 0x0000000000000FF0;
GreenMask: 0x0000000000000FF0;
BlueMask: 0x0FF0000000000000;

//Started: 221105 09:02:39
//Finished: 221105 09:26:14
```

Table C.4: EVVR input data
C.3 Simulations maintaining multiple genotypes

This Appendix gives fuller details of the simulations described in Section 5.3 on page 96.

C.3.1 Genotype diversity with zygote testing

All these cases were run with their initial populations set up in three configurations:

(i) *Uniform Random Comp* (URC) where every plant has the same, randomly selected, complementary pair;

(ii) *Uniform Random Zero* (URZ) where every plant has the same random chromosome paired with a zero chromosome; and

(iii) *Random Zero* (RZ) where each cell has a different random chromosome paired with a zero.

The recombination probability ($P_R$) was set to 0.2, 0.4 and 0.8 and all scenarios were rerun 20 times with different sequences of random numbers to provide an average outcome in each case. No additional mutations were added during reproduction (i.e. $P_m = 0$) so that all genetic changes were due to recombination.

The output of genotype diversity was provided by the type-F report which reported populations by genotype.

These simulations are discussed in Section 5.3.1 on page 97.

C.3.1.1 Genotype diversity with zygote testing: input

The scenarios (FZG0) were run for 15000 time-steps. In order to check the long-term effects, the URC case (FZG1) with $P_R = 0.4$ was run for 50000 time-steps and (FZG2) with $P_R = 0.2$ for 150000 time-steps.

The input data files are shown in Tables C.1 to C.3.
Figure C.1: Zygotic case FZG0 input data
Figure C.2: Zygotic case FZG1: URC, $P_R = 0.4$ and 50000 time-steps
Figure C.3: Zygotic case FZG2: URC, $P_R = 0.2$ and 150000 time-steps
C.3.1.2 Genotype diversity with zygote testing: results

The runs FZG0-000 to -002 started from the “URC” initial population, using $P_R$ of 0.2, 0.4 and 0.8.

Figure C.4: Zygotic case FZG0 (URC): growth in diversity
The runs FZG0-003 to 005 started from the “URZ” initial population.

Figure C.5: Zygotic case FZG0 (URZ): growth in diversity
The runs FZG0-006 to 007 started from the “RZ” initial population.

Figure C.6: Zygotic case FZG0 (RZ): growth in diversity
The FZG1 and FZG2 runs started from the “URC” initial population with $P_R$ 0.4 and 0.2 respectively.

Figure C.7: Zygotic case FZG1 & FZG2: growth in diversity
C.3.2 Preliminary test of phenotype matching (PT03)

These simulations are discussed in Section 5.3.3 on page 98.

C.3.2.1 Preliminary test of phenotype matching (PT03): input

The input data for this simulation is shown in Table C.8 on page 192.

```plaintext
//TITLES
Computer: Laptop;
Purpose: Phenotype Similarity with Tolerance Long run;
Prefix: PT03;
OutFolder: ..

//LATTICE
Rows: 32;
Cols: 128;
WrapRows: no;
WrapCols: YES;
SwapsPerCycle: 0.25;

//INITIAL SETUP
InitialSet: 0x0F0F0F0F0F0F0F0F 0x0000000000000000;

//BASE DATA
Mate: 1;
Baby: 1;
BirthAge: 1;
BirthLoc: 0;
LociPerYr: 0.8;
Mutation: 0;
Mutation: 1.0;
Recomb: 0.70;
Recomb: 0.00;
Phenotypeloci: 0xFFFFFFFFFFFFFFFF; //governs compatibility
Phenotypeloc: 8;

//ASSORTATIVE MATING
AssocativeLoc: 0x0000000000000000;

//PREDICTORS (Preditor death rate is added to EDR per mutation)
PredProtectLoc: 0x0000000000000000;
PredExposerLoc: 0x0000000000000000;
PredperProtectMutation: 0.000;
PredperExposerMutation: 0.000;

//ENV DEATH
EnvDeathRate: 0.040;

//CYCLES
Generations: 800000;
Steps: 20; //excluding the report at time 0
GeometricSteps: No;
Seed: 9;
Reruns: 10;

//REPORTING
ReportDoubleMutations: YES;  //rptMd
GenePlotInterval: 3;

//IMAGES
ImageHeight: 900; //max in pixels
ImageWidth: 1600; //max in pixels
CaptureReruns: 2; //images of reruns

//Colour mapping
GenImage: YES;
RedMask: 0x00000000FF000000;
GreenMask: 0x00000000FF000000;
BlueMask: 0x00000000FF000000;

//RUNTIME
//VERSION: Sympatria 2.1
//Started: 131121 22:47:07
//Finished: 131122 06:16:49
```

Figure C.8: Preliminary test of phenotype matching (PT03): input
C.3.2.2 Preliminary test of phenotype matching (PT03): results

Figure C.9 on page 193 shows the longterm stability in a complex gene-patterns achieved by using phenotype matching.

Figure C.9: Longterm stable gene patterns (FPT0): results
C.3.3 Phenotype matching with tolerance of difference (FPT0)

These simulations are discussed in Section 5.3.4 on page 99.

C.3.3.1 Phenotype matching with tolerance of difference (FPT0): input

Cases FPT0 used increasing tolerance of phenotype difference ($T_{pm}$). The 45 input combinations for FPT0 are summarised in Table C.5 on page 194. In all cases the probability of additional mutation ($P_m$) was zero and each combination was rerun 20 times.

<table>
<thead>
<tr>
<th>Initial: UR C</th>
<th>Initial: URZ</th>
<th>Initial: RZ</th>
<th>Recombination $P_R$</th>
<th>Mutation $P_m$</th>
<th>PM Tolerance $T_{pm}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ref FPT0-000</td>
<td>Ref FPT0-015</td>
<td>Ref FPT0-030</td>
<td>20%</td>
<td>0%</td>
<td>0</td>
</tr>
<tr>
<td>001</td>
<td>016</td>
<td>031</td>
<td>20%</td>
<td>0%</td>
<td>1</td>
</tr>
<tr>
<td>002</td>
<td>017</td>
<td>032</td>
<td>20%</td>
<td>0%</td>
<td>2</td>
</tr>
<tr>
<td>003</td>
<td>018</td>
<td>033</td>
<td>20%</td>
<td>0%</td>
<td>4</td>
</tr>
<tr>
<td>004</td>
<td>019</td>
<td>034</td>
<td>20%</td>
<td>0%</td>
<td>8</td>
</tr>
<tr>
<td>005</td>
<td>020</td>
<td>035</td>
<td>40%</td>
<td>0%</td>
<td>0</td>
</tr>
<tr>
<td>006</td>
<td>021</td>
<td>036</td>
<td>40%</td>
<td>0%</td>
<td>1</td>
</tr>
<tr>
<td>007</td>
<td>022</td>
<td>037</td>
<td>40%</td>
<td>0%</td>
<td>2</td>
</tr>
<tr>
<td>008</td>
<td>023</td>
<td>038</td>
<td>40%</td>
<td>0%</td>
<td>4</td>
</tr>
<tr>
<td>009</td>
<td>024</td>
<td>039</td>
<td>40%</td>
<td>0%</td>
<td>8</td>
</tr>
<tr>
<td>010</td>
<td>025</td>
<td>040</td>
<td>80%</td>
<td>0%</td>
<td>0</td>
</tr>
<tr>
<td>011</td>
<td>026</td>
<td>041</td>
<td>80%</td>
<td>0%</td>
<td>1</td>
</tr>
<tr>
<td>012</td>
<td>027</td>
<td>042</td>
<td>80%</td>
<td>0%</td>
<td>2</td>
</tr>
<tr>
<td>013</td>
<td>028</td>
<td>043</td>
<td>80%</td>
<td>0%</td>
<td>4</td>
</tr>
<tr>
<td>014</td>
<td>029</td>
<td>044</td>
<td>80%</td>
<td>0%</td>
<td>8</td>
</tr>
</tbody>
</table>

Table C.5: Summary of assumptions for FPT0 cases

The initial population configurations are:

URC: Uniform Random Complement;
URZ: Uniform Random Zero;
RZ: Random Zero.

These initial populations are fully described in Appendix A.1.1.1 on page 158.

The input data for these simulations is shown in Table C.6 on page 195.
Purpose: Family numbers. Phenotype matching varying. No mutation;
Prefix: FPT0;

Rows: 50; //df 3
Cols: 50; //df 3
WrapRows: No; //df No
WrapCols: No; //df No
SwapsPerCycle: 0.10; //df 0.10

InitialSet: UniformRandomComp, UniformRandomZero, RandomZero;

MateRange: 1; //df 1
SeedRange: 1; //df 1
MatureAge: 1; //df 1
Seeds: 1; //df 1
Recomb: 0.2, 0.4, 0.8; //df 0.40
CrossOvers: 1.000; //df 1;
Mutation: 0.000; //df 0.001

PhenotypeLoci: 0xFFFFFFFFFFFFFFFF; //df ZERO implies not used
PhenoTolerance: 0, 1, 2, 4, 8; //df 0

AssortativeLoci: 0x0000000000000000; //df ZERO implies not used
AssortativeFactor: 0; //increases influence of AM on mate selection. df 0

Generations: 15000; //df 20
Stops: 20; //excluding the report at time 0 //df 20
RandomSeed: 99; //df 0 (deprecated form 'Seed')

ReportFamilies: YES; //df No
ReportLoci: 0xFFFFFFFFFFFFFFFF; /df 0xFFFFFFFFFFFFFFFF

GeneImages: YES; //df No

CaptureReruns: 1; //Images of reruns. df 1

RedMask: 0xFF00000000000000; //df 0xFF0000
GreenMask: 0x0000000FF0000000; //df 0x00FF00
BlueMask: 0x00000000000000FF; //df 0x0000FF

MakeMovie: YES; //df No
ImageDwell: 1.0; //seconds per frame. df 1.0
KeepImages: YES; //keep image files after compiling movie. df Yes

Table C.6: Phenotype matching with varying tolerance (FPT0) input data
C.3.3.2 Phenotype matching with tolerance of difference (FPT0): results

Figure C.10 on page 197 and Figure C.11 on page 198 show a decline after some initial genetic diversity starting from URC and URZ populations respectively. Only when tolerance of phenotypic difference ($T_{pm}$) exceed 4 does the diversity increase continuously.

Figure C.12 on page 199 shows the rapid decline in genetic diversity starting from a RZ population which has 2500 different random genotypes which die out completely.
Figure C.10: Change in genetic diversity. URC, $P_R = 20\%\ 40\%\ 80\%\ (FPT0)$
Figure C.11: Change in genetic diversity. URZ, $P_R = 20\% 40\% 80\%$ (FPT0)
Figure C.12: Change in genetic diversity. \( RZ, P_R = 20\% 40\% 80\% \) (FPT0)
C.3.4 Phenotype matching with various mutation rates (FPT1)

These simulations are discussed in Section 5.3.5 on page 104.

C.3.4.1 Phenotype matching with various mutation rates (FPT1): input

Cases FPT1 used zero tolerance of phenotype difference \( (T_{pm}) \) but increasing incidence of additional mutation \( (P_m) \). The 36 input combinations for FPT1 are summarised in Table C.7 on page 200. In all cases the probability of additional mutation \( (P_m) \) was zero and each combination was rerun 20 times.

<table>
<thead>
<tr>
<th>Initial: URC Ref FPT1-</th>
<th>Initial: URZ Ref: FPT1-</th>
<th>Initial: RZ Ref: FPT1-</th>
<th>Recombination ( P_R )</th>
<th>Mutation ( P_m )</th>
<th>PM Tolerance ( T_{pm} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>000</td>
<td>012</td>
<td>024</td>
<td>20%</td>
<td>1%</td>
<td>0</td>
</tr>
<tr>
<td>001</td>
<td>013</td>
<td>025</td>
<td>20%</td>
<td>10%</td>
<td>0</td>
</tr>
<tr>
<td>002</td>
<td>014</td>
<td>026</td>
<td>20%</td>
<td>50%</td>
<td>0</td>
</tr>
<tr>
<td>003</td>
<td>015</td>
<td>027</td>
<td>20%</td>
<td>100%</td>
<td>0</td>
</tr>
<tr>
<td>004</td>
<td>016</td>
<td>028</td>
<td>40%</td>
<td>1%</td>
<td>0</td>
</tr>
<tr>
<td>005</td>
<td>017</td>
<td>029</td>
<td>40%</td>
<td>10%</td>
<td>0</td>
</tr>
<tr>
<td>006</td>
<td>018</td>
<td>030</td>
<td>40%</td>
<td>50%</td>
<td>0</td>
</tr>
<tr>
<td>007</td>
<td>019</td>
<td>031</td>
<td>40%</td>
<td>100%</td>
<td>0</td>
</tr>
<tr>
<td>008</td>
<td>020</td>
<td>032</td>
<td>80%</td>
<td>1%</td>
<td>0</td>
</tr>
<tr>
<td>009</td>
<td>021</td>
<td>033</td>
<td>80%</td>
<td>10%</td>
<td>0</td>
</tr>
<tr>
<td>010</td>
<td>022</td>
<td>034</td>
<td>80%</td>
<td>50%</td>
<td>0</td>
</tr>
<tr>
<td>011</td>
<td>023</td>
<td>035</td>
<td>80%</td>
<td>100%</td>
<td>0</td>
</tr>
</tbody>
</table>

Table C.7: Summary of assumptions for FPT1 cases

The initial population configurations are:
URC: Uniform Random Complement;
URZ: Uniform Random Zero;
RZ: Random Zero.
These initial populations are fully described in Appendix A.1.1.1 on page 158.

The input data for these simulations is shown in Table C.8 on page 201.

C.3.4.2 Phenotype matching with various mutation rates (FPT1): results

Figure C.13 on page 203 shows an initial rise in diversity which then stabilises to a few hundred genotypes. In Figure C.14 on page 204 with UniformRandomZero (URZ) initial population the level of diversity appear to stabilise at slightly lower levels than in the URC population but in both cases the ultimate level of diversity seems insensitive to the recombination probability...
Table C.8: Phenotype matching with varying mutation (FPT1) input data
(P_R) but the initial peak diversity is larger for lower values of P_R. Diversity stabilises in all cases except with the highest level of additional mutation.

Figure C.15 on page 205 shows that, as with case FPT0, the random starting population begins with total diversity (2500 random genotypes) but no types become established colonies.
Figure C.13: Change in genetic diversity. URC, $P_R = 20\% 40\% 80\%$ (FPT1)
Figure C.14: Change in genetic diversity. URZ, $P_R = 20\%$ 40\% 80\% (FPT1)
Figure C.15: Change in genetic diversity. RZ, $P_R = 20\%$ 40\% 80\% (FPT1)
C.3.5 Simulations with only assortative mating (FAM0)

These simulations are discussed in Section 5.3.6 on page 107.

C.3.5.1 Simulations with only assortative mating (FAM0): input

Table C.9 on page 207 shows the input data for the FAM0 simulations of assortative mating.

C.3.5.2 Simulations with only assortative mating (FAM0): results

Figures C.16 to C.18 show the diversity of genotypes for each of the 27 cases tested. Each page shows three graph for $P_R = 20\%, 40\% & 60\%$. There is a page for the graphs for each of the three initial population URC, URZ and RZ.
Table C.9: Assortative mating with different $P_R$ and $g_{am}$ (FAM0) input data
Figure C.16: Change in genetic diversity. $URC, P_R = 20\% \ 40\% \ 80\%$ (FAM0)
Figure C.17: Change in genetic diversity. URZ, $P_R = 20\% 40\% 80\%$ (FAM0)
Figure C.18: Change in genetic diversity. RZ, $P_R = 20\% \ 40\% \ 80\%$ (FAM0)
C.3.6  PM and AM with initial URZ (PTAM5)

These simulations are discussed in Section 5.3.7 on page 110.

C.3.6.1  PM and AM with initial URZ (PTAM5): input

```plaintext
//Sympatrria input file
Computer: Toshiba Satellite T110-107 Laptop;
//TITLES
Purpose: Close-linked phenotype & assortative selection;
//LATTICE
Rows: 50; //df 3
Cols: 50; //df 3
WrapRows: No; //df No
WrapCols: No; //df No
//INITIAL SETUP
InitialSet: UniformRandomZero;
//BASE DATA
MateRange: 8; //df 1
SeedRange: 2; //df 1
MatureAge: 1; //df 1
Seeds: 1; //df 1
Recomb: 0.6; //df 0.48
CrossOver: 1.000; //df 1;
Mutation: 0.001; //df 0.001
EnvDeathRate: 0.01; //df 0.01
//PHENOTYPE COMPATIBILITY
PhenotypeLoci: 0x000000FA5F000000; //df ZERO implies not used
PhenoTolerance: 0; //df 0
//ASSORTATIVE MATING
AssortativeLoci: 0x00000005A0000000; //df ZERO implies not used
AssortativeIndex: 4;
//CYCLES
Generations: 20000; //df 20
Stops: 20; //excluding the report at time 0 //df 20
ReRuns: 10; //df 1
RandomSeed: 99; //df 0 (deprecated form 'Seed')
//REPORTING
//rptF: Nr of Family types v Time
ReportFamilies: YES; //df No
ReportLoci: 0x000000FFFF000000; //df 0xFFFFFFFFFFFFFFFF
//IMAGE ATTRIBUTES
CompatibilityImages: No; //df No
ColonyImages: No; //df No
GeneImages: YES; //df No
//Size
ImageHeight: 900; //max in pixels. df 900
ImageWidth: 1600; //max in pixels. df 1600
CaptureReruns: 1; //Images of reruns. df 1
//Colour mapping
RedMask: 0xF000000000000000; //df 0xFF0000
GreenMask: 0x00F0000000000000; //df 0x00FF00
BlueMask: 0x00000000000000FF; //df 0x0000FF
//Movie maker
MakeMovie: YES; //df No
ImageDwell: 1.0; //seconds per frame. df 1.0
KeepImages: YES; //keep image files after compiling movie. df Yes
```

Table C.10: PM with AM and initial URZ (PTAM0) input data
C.3.6.2 PM and AM with initial URZ (PTAM5): results

Figure C.19: Genetic diversity over time (PTAM5)
C.3.7 De-linked PM and AM with two colonies (DLPA0)

These simulations are discussed in Section 5.3.8 on page 115.

C.3.7.1 De-linked PM and AM with two colonies (DLPA0): input

```
//Sympatria input file
Computer: Toshiba Satellite T110-107 Laptop;
//TITLES
Purpose: De-linked AM selection & PT matching;
//LATTICE
Rows: 50; //df 3
Cols: 50; //df 3
WrapRows: No; //df No
WrapCols: No; //df No

//INITIAL SETUP
InitialSet: LeftRight FF00000000000000_0000000000000000 00000000000000FF_0000000000000000,
LeftRight FF00000000000000_FF00000000000000 00000000000000FF_00000000000000FF;

//BASE DATA
MateRange: 4; //df 1
SeedRange: 2; //df 1
MatureAge: 1; //df 1
Seeds: 1; //df 1
Recomb: 0.6, 1.0; //df 0.40
CrossOvers: 1.000; //df 1;
Mutation: 0.000; //df 0.001

//PHENOTYPE COMPATIBILITY
PhenotypeLoci: 0xFF00000000000000; //df ZERO implies not used
PhenoTolerance: 0,2; //df 0

//ASSORTATIVE MATING
AssortativeLoci: 0x00000000000000FF; //df ZERO implies not used
AssortativeIndex: 4;

//CYCLES
Generations: 20000; //df 20
Stops: 20; //excluding the report at time 0 //df 20
ReRuns: 18; //df 1
RandomSeed: 99; //df 0 (deprecated form 'Seed')

//REPORTING
//rptF: Nr of Family types v Time
ReportFamilies: Yes; //df No
ReportLoci: 0xFFFFFFFFFFFFFFFF; //df 0xFFFFFFFFFFFFFFFF

//IMAGE ATTRIBUTES
CompatibilityImages: No; //df No
ColonyImages: No; //df No
GeneImages: Yes; //df No

//Size
ImageHeight: 900; //max in pixels. df 900
ImageWidth: 1600; //max in pixels. df 1600
CaptureReruns: 1; //Images of reruns. df 1

//Colour mapping
RedMask: 0xFF00000000000000; //df 0xFF0000
GreenMask: 0x00FF000000000000; //df 0x00FF00
BlueMask: 0x0000FF0000000000; //df 0x0000FF

//Movie maker
MakeMovie: Yes; //df No
ImageDwell: 0.2; //seconds per frame. df 1.0
KeepImages: Yes; //keep image files after compiling movie. df Yes
```

Table C.11: Delinked PM and AM with 2 colonies (DLPA0) input data
C.3.7.2 De-linked PM and AM with two colonies (DLPA0): results

The percentage populations shown in Figure C.20 on page 215 are the average population of each genotype for 10 reruns of each simulation.

Table C.12 shows the diversity (as numbers of distinct genotypes) averaged over 10 reruns for various values of probability of recombination ($P_R$) and pehotype matching tolerance ($T_{pm}$).

<table>
<thead>
<tr>
<th>Run Ref.</th>
<th>DLPA0-000</th>
<th>001</th>
<th>002</th>
<th>003</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recombination</td>
<td>$P_R$</td>
<td>0.60</td>
<td>0.60</td>
<td>1.00</td>
</tr>
<tr>
<td>PM tolerance</td>
<td>$T_{pm}$</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Time-step</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1000</td>
<td></td>
<td>87</td>
<td>119</td>
<td>145</td>
</tr>
<tr>
<td>2000</td>
<td></td>
<td>48</td>
<td>92</td>
<td>89</td>
</tr>
<tr>
<td>3000</td>
<td></td>
<td>32</td>
<td>69</td>
<td>62</td>
</tr>
<tr>
<td>4000</td>
<td></td>
<td>22</td>
<td>55</td>
<td>41</td>
</tr>
<tr>
<td>5000</td>
<td></td>
<td>17</td>
<td>44</td>
<td>33</td>
</tr>
<tr>
<td>6000</td>
<td></td>
<td>13</td>
<td>38</td>
<td>24</td>
</tr>
<tr>
<td>7000</td>
<td></td>
<td>9</td>
<td>32</td>
<td>19</td>
</tr>
<tr>
<td>8000</td>
<td></td>
<td>9</td>
<td>29</td>
<td>19</td>
</tr>
<tr>
<td>9000</td>
<td></td>
<td>7</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>10000</td>
<td></td>
<td>6</td>
<td>19</td>
<td>16</td>
</tr>
<tr>
<td>11000</td>
<td></td>
<td>5</td>
<td>20</td>
<td>18</td>
</tr>
<tr>
<td>12000</td>
<td></td>
<td>5</td>
<td>19</td>
<td>13</td>
</tr>
<tr>
<td>13000</td>
<td></td>
<td>5</td>
<td>17</td>
<td>12</td>
</tr>
<tr>
<td>14000</td>
<td></td>
<td>4</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>15000</td>
<td></td>
<td>4</td>
<td>15</td>
<td>9</td>
</tr>
<tr>
<td>16000</td>
<td></td>
<td>5</td>
<td>15</td>
<td>7</td>
</tr>
<tr>
<td>17000</td>
<td></td>
<td>5</td>
<td>14</td>
<td>6</td>
</tr>
<tr>
<td>18000</td>
<td></td>
<td>5</td>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td>19000</td>
<td></td>
<td>4</td>
<td>16</td>
<td>5</td>
</tr>
<tr>
<td>20000</td>
<td></td>
<td>4</td>
<td>15</td>
<td>5</td>
</tr>
</tbody>
</table>

Table C.12: Average diversity all with AM $q_{am} = 4$ (DLPA0)
Figure C.20: Average genotypes population $P_R = 60\%$ and $q_{am} = 4$ (DLPA0-000 & -001)
Figure C.21 on page 216 shows a typical distribution of genotypes after 20000 time-steps when no additional mutations have been introduced.

In Table C.13 on page 217 the genotype are classified as $A$, $B$ or $0$ to indicate their relationship to the initial genotypes $A0$ and $B0$, while $x$ indicates that a different chromosome is present which is not identical to $A$, $B$ or $0$. 
<table>
<thead>
<tr>
<th>Pop% at t=20000</th>
<th>Chromosomes</th>
<th>Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>44.000%</td>
<td>0a0000 0000 0000 0000 &amp; 0a0000 0000 0000 0000</td>
<td>00</td>
</tr>
<tr>
<td>0.108%</td>
<td>0a0000 0000 0000 00FF &amp; 0a0000 0000 0000 0000</td>
<td>A0</td>
</tr>
<tr>
<td>28.228%</td>
<td>0a0000 0000 0000 00FF &amp; 0a0000 0000 0000 00FF</td>
<td>AA</td>
</tr>
<tr>
<td>0.000%</td>
<td>0xF000 0000 0000 0000 &amp; 0x0000 0000 0000 00FF</td>
<td>BA</td>
</tr>
<tr>
<td>0.000%</td>
<td>0xF000 0000 0000 0000 &amp; 0x0000 0000 0000 0000</td>
<td>B0</td>
</tr>
<tr>
<td>22.328%</td>
<td>0xF000 0000 0000 0000 &amp; 0xF000 0000 0000 0000</td>
<td>BB</td>
</tr>
<tr>
<td>2.844%</td>
<td>0xF000 0000 0000 0000 &amp; 0xF000 0000 0000 0000</td>
<td>xx</td>
</tr>
<tr>
<td>1.136%</td>
<td>0xE000 0000 0000 0000 &amp; 0xE000 0000 0000 0000</td>
<td>xx</td>
</tr>
<tr>
<td>0.212%</td>
<td>0xF000 0000 0000 0000 &amp; 0xF000 0000 0000 00FF</td>
<td>xx</td>
</tr>
<tr>
<td>0.076%</td>
<td>0x0000 0000 0000 00FF &amp; 0x0000 0000 0000 00FF</td>
<td>Ax</td>
</tr>
<tr>
<td>0.056%</td>
<td>0x0000 0000 0000 00FF &amp; 0x0000 0000 0000 00FF</td>
<td>Ax</td>
</tr>
<tr>
<td>0.036%</td>
<td>0x0000 0000 0000 00FF &amp; 0x0000 0000 0000 00FF</td>
<td>Ax</td>
</tr>
<tr>
<td>0.020%</td>
<td>0x0000 0000 0000 00FF &amp; 0x0000 0000 0000 00FF</td>
<td>Ax</td>
</tr>
<tr>
<td>0.016%</td>
<td>0x0000 0000 0000 00FF &amp; 0x0000 0000 0000 00FF</td>
<td>Ax</td>
</tr>
<tr>
<td>0.016%</td>
<td>0x0000 0000 0000 00FF &amp; 0x0000 0000 0000 00FF</td>
<td>Ax</td>
</tr>
<tr>
<td>0.012%</td>
<td>0x0000 0000 0000 00FF &amp; 0x0000 0000 0000 00FF</td>
<td>Ax</td>
</tr>
<tr>
<td>0.012%</td>
<td>0x0000 0000 0000 00FF &amp; 0x0000 0000 0000 00FF</td>
<td>Ax</td>
</tr>
<tr>
<td>0.008%</td>
<td>0x0000 0000 0000 00FF &amp; 0x0000 0000 0000 00FF</td>
<td>Ax</td>
</tr>
<tr>
<td>0.008%</td>
<td>0x0000 0000 0000 00FF &amp; 0x0000 0000 0000 00FF</td>
<td>Ax</td>
</tr>
<tr>
<td>0.004%</td>
<td>0x0000 0000 0000 00FF &amp; 0x0000 0000 0000 00FF</td>
<td>Ax</td>
</tr>
<tr>
<td>0.004%</td>
<td>0x0000 0000 0000 00FF &amp; 0x0000 0000 0000 00FF</td>
<td>Ax</td>
</tr>
<tr>
<td>0.004%</td>
<td>0x0000 0000 0000 00FF &amp; 0x0000 0000 0000 00FF</td>
<td>x0</td>
</tr>
<tr>
<td>0.004%</td>
<td>0x0000 0000 0000 00FF &amp; 0x0000 0000 0000 00FF</td>
<td>x0</td>
</tr>
</tbody>
</table>

Table C.13: Populations of genotypes at t=20000 (DLPA0-000) with $T_{pm} = 0$
C.3.8 Unconstrained diversity

This simulation is discussed in Section 5.3.9 on page 120.

C.3.8.1 Unconstrained diversity: input

```plaintext
//Sympatria input file
Computer: Toshiba Satellite T110-107 Laptop;

//TITLES
Purpose: Family numbers. No genetic preferences. No mutation;
Prefix: FUN0;

//LATTICE
Rows: 50; //df 3
Cols: 50; //df 3
WrapRows: No; //df No
WrapCols: No; //df No
SwapsPerCycle: 0.10; //df 0.10

//INITIAL SETUP
InitialSet: UniformRandomComp, UniformRandomZero, RandomZero;

//BASE DATA
MateRange: 1; //df 1
SeedRange: 1; //df 1
MatureAge: 1; //df 1
Seeds: 1; //df 1
Recomb: 0.2, 0.4, 0.8; //df 0.40
CrossOvers: 1.00; //df 1;
Mutation: 0.000; //df 0.001

//AGEING (df FALSE)
BirthLoci: 0; //df 0
LociPerYr: 0.0; //df 0.0
MutTolerance: 64; //df 64

//PHENOTYPE COMPATIBILITY
PhenotypeLoci: 0x0000000000000000; //df ZERO implies not used
PhenoTolerance: 0; //df 0

//ASSORTATIVE MATING
AssortativeLoci: 0x0000000000000000; //df ZERO implies not used
AssortativeFactor: 0; //increases influence of AM on mate selection. df 0

//CYCLES
Generations: 15000; //df 20
Stops: 20; //excluding the report at time 0 //df 20
ReRuns: 20; //df 1
RandomSeed: 99; //df 0 (deprecated form 'Seed')

//REPORTING
//rptF: Nr of Family types v Time
ReportFamilies: YES; //df No
ReportLoci: 0xFFFFFFFFFFFFFFFF; //df 0xFFFFFFFFFFFFFFFF

//IMAGE ATTRIBUTES
GeneImages: YES; //df No
CaptureReruns: 1; //Images of reruns. df 1

//Colour mapping
RedMask: 0xFF00000000000000; //df 0xFF0000
GreenMask: 0x0000000FF00000000; //df 0x00FF00
BlueMask: 0x000000000000000FF; //df 0x0000FF
MakeMovie: YES; //df No
ImageDwell: 1.0; //seconds per frame. df 1.0
KeepImages: YES; //keep image files after compiling movie. df Yes
```

Table C.14: Evolution with no constraints (FUN0) input data

C.3.8.2 Unconstrained diversity: results
Figure C.22: Genetic diversity from initial URC pop. (FUN0)
Figure C.23: Genetic diversity from initial URZ pop. (FUN0)
Figure C.24: Genetic diversity from initial RZ pop. (FUN0)
C.4 Simulations with the Dobzhansky-Müller mechanism

This Appendix gives fuller details of the simulations using the *Sympatria* model to investigate aspects of the Dobzhansky-Müller mechanism. These simulations are described in Section 5.4.

C.4.1 Initial Dobzhansky-Müller simulation (DM04)

This was the first simulation to simulate the Dobzhansky-Müller model.

C.4.1.1 Initial Dobzhansky-Müller simulation (DM04): input

The key changes between each run were to the ‘PhenotypeLoci’ and the ‘AssortativeLoci’ which were alternately set to zero and 0x3. The input file for this simulation is shown in Table C.15 on page 223.

C.4.1.2 Initial Dobzhansky-Müller simulation (DM04): results

The following data shows the results of modelling the Dobzhansky-Müller model using and not using phenotype matching and assortative mating. The results are shown in Figures C.25 to C.32. In each case the first figure shows the percentage of the population with each genotype and the second shows the development of genotype groups as a map of the lattice.

The two chromosomes in each plant are distinguished on the gene map by using the input option 'HapCombination=SPLIT'. This ensures that the single and double mutations appear as different colours in the gene mapping.

Report ‘F’ was specified to give details of the populations of different genotypes.

The population development and the genotype mapping are shown for each of the four combinations of input options. In each case the average population distribution is shown as a histogramme and the average number of distinct genotypes present as a line-graph over time.

The genotype maps are shown every 5,000 time-steps starting from 5,000. They are set to distinguish between plants with single mutations (a darker shade) than those with a double mutation. The initial unmuted form (aabb) appear as blue; the Aabb and AAbb form appear in two shades of turquoise; and the aaBb and aaBB forms appear in two shades of purple.

It should be noted that the genotype mapping reproduced here show only one of the 20 times that the example was run. The development of the population in that particular case will not therefore exactly match the averages shown in the histogramme.

The results are discussed in Section 5.4.1.2.
//Sympatria input file
Computer: Toshiba Satellite T110-107 Laptop;

//TITLES
Purpose: Dobzhansky-Muller with and without PM and AM;

//LATTICE
Rows: 50;
Cols: 50;

//INITIAL SETUP
InitialSet: 0xFFFFFFFFFFFFFFFC 0xFFFFFFFFFFFFFFFC;

//BASE DATA
MateRange: 10;
SeedRange: 1;
MatureAge: 1;
Seeds: 1;
Recomb: 0.80;
CrossOvers: 1.00;
Mutation: 1.00;

//PHENOTYPE FATALITY (a fatal combination of mutations)
PhenotypeFatal: 0x0000000000000003;

//PHENOTYPE COMPATIBILITY
PhenotypeLoci: 0x0000000000000000, 0x0000000000000003;
PhenoTolerance: 1;

//ASSORTATIVE MATING
AssortativeLoci: 0x0000000000000000, 0x0000000000000003;
AssortativeIndex: 8;

//CYCLES
Generations: 100000;
Stops: 20;
ReRuns: 20;
RandomSeed: 999;

//rptF: Nr of Genotypes v Time
ReportFamilies: YES;
ReportLoci: 0xFFFFFFFFFFFFFFF;

//Colour mapping
GeneImages: YES; //df No
CaptureReruns: 1;
RedMask: 0x0000000000000001;
GreenMask: 0x0000000000000002;
BlueMask: 0x0000000000000004;
HapCombination: SPLIT; //df HI Options HI, LO, OR, AND, SPLIT
### Analysis of populations by genotype

**Source:** DM04_reordFOOD.csv

**Max prop.** 2500  
**Reruns** 20

#### Total populations of genotypes for all reruns

<table>
<thead>
<tr>
<th>Genotype</th>
<th>% of maximum population</th>
</tr>
</thead>
<tbody>
<tr>
<td>aabb</td>
<td>100.0%</td>
</tr>
<tr>
<td>Aabb</td>
<td>99.1%</td>
</tr>
<tr>
<td>aA bb</td>
<td>99.1%</td>
</tr>
<tr>
<td>AA bb</td>
<td>99.1%</td>
</tr>
<tr>
<td>aabb</td>
<td>99.0%</td>
</tr>
<tr>
<td>Aabb</td>
<td>99.1%</td>
</tr>
<tr>
<td>aAbb</td>
<td>99.1%</td>
</tr>
<tr>
<td>AA bb</td>
<td>99.1%</td>
</tr>
</tbody>
</table>

#### Instances of genotypes

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Average no. of genotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>aabb</td>
<td>2500</td>
</tr>
<tr>
<td>Aabb</td>
<td>2500</td>
</tr>
<tr>
<td>aAbb</td>
<td>2500</td>
</tr>
<tr>
<td>AA bb</td>
<td>2500</td>
</tr>
</tbody>
</table>

**Case:** PhenoTypeLoc=0x0000000000000000; AssortativeLoc=0x0000000000000000;

---

**Figure C.25:** DM04-000 Population data

---

---
Figure C.26: DM04-000(000) Example genotype maps at 5k time-steps
Figure C.27: DM04-001 Population data
Figure C.28: DM04-001(000) Example genotype maps at 5k time-steps
### Analysis of populations by genotype

**Source:** DM04_reportF002.csv  
**Max pop.:** 2500  
**Reruns:** 20

#### Total populations of genotypes for all reruns

<table>
<thead>
<tr>
<th>Alleles</th>
<th>0</th>
<th>5000</th>
<th>10000</th>
<th>15000</th>
<th>20000</th>
<th>25000</th>
<th>30000</th>
<th>35000</th>
<th>40000</th>
<th>45000</th>
<th>50000</th>
<th>55000</th>
<th>60000</th>
<th>65000</th>
<th>70000</th>
<th>75000</th>
<th>80000</th>
<th>85000</th>
<th>90000</th>
<th>95000</th>
<th>100000</th>
</tr>
</thead>
<tbody>
<tr>
<td>aabb</td>
<td>0</td>
<td>5000</td>
<td>10000</td>
<td>15000</td>
<td>20000</td>
<td>25000</td>
<td>30000</td>
<td>35000</td>
<td>40000</td>
<td>45000</td>
<td>50000</td>
<td>55000</td>
<td>60000</td>
<td>65000</td>
<td>70000</td>
<td>75000</td>
<td>80000</td>
<td>85000</td>
<td>90000</td>
<td>95000</td>
<td>100000</td>
</tr>
<tr>
<td>Aabb</td>
<td>0</td>
<td>5000</td>
<td>10000</td>
<td>15000</td>
<td>20000</td>
<td>25000</td>
<td>30000</td>
<td>35000</td>
<td>40000</td>
<td>45000</td>
<td>50000</td>
<td>55000</td>
<td>60000</td>
<td>65000</td>
<td>70000</td>
<td>75000</td>
<td>80000</td>
<td>85000</td>
<td>90000</td>
<td>95000</td>
<td>100000</td>
</tr>
<tr>
<td>aAabb</td>
<td>0</td>
<td>5000</td>
<td>10000</td>
<td>15000</td>
<td>20000</td>
<td>25000</td>
<td>30000</td>
<td>35000</td>
<td>40000</td>
<td>45000</td>
<td>50000</td>
<td>55000</td>
<td>60000</td>
<td>65000</td>
<td>70000</td>
<td>75000</td>
<td>80000</td>
<td>85000</td>
<td>90000</td>
<td>95000</td>
<td>100000</td>
</tr>
<tr>
<td>aAABb</td>
<td>0</td>
<td>5000</td>
<td>10000</td>
<td>15000</td>
<td>20000</td>
<td>25000</td>
<td>30000</td>
<td>35000</td>
<td>40000</td>
<td>45000</td>
<td>50000</td>
<td>55000</td>
<td>60000</td>
<td>65000</td>
<td>70000</td>
<td>75000</td>
<td>80000</td>
<td>85000</td>
<td>90000</td>
<td>95000</td>
<td>100000</td>
</tr>
</tbody>
</table>

#### Average for all reruns of percentage of maximum population for each genotype

| Alleles | Time | 0 | 5000 | 10000 | 15000 | 20000 | 25000 | 30000 | 35000 | 40000 | 45000 | 50000 | 55000 | 60000 | 65000 | 70000 | 75000 | 80000 | 85000 | 90000 | 95000 | 100000 |
|---------|------|---|------|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| aabb    | 100.0% | 64.7% | 62.2% | 60.7% | 58.6% | 56.5% | 52.8% | 47.2% | 44.3% | 37.1% | 35.5% | 28.7% | 25.3% | 22.6% | 17.0% | 14.4% | 10.5% | 8.6% | 5.9% | 4.8% | 6.8% |
| Aabb    | 0.0% | 14.2% | 14.4% | 13.9% | 13.4% | 13.4% | 12.8% | 12.3% | 12.4% | 11.7% | 11.4% | 11.4% | 11.0% | 10.6% | 9.8% | 9.8% | 9.1% | 9.1% | 8.8% | 8.7% | 7.9% | 7.4% |
| aAabb   | 0.0% | 17.4% | 17.5% | 10.1% | 21.0% | 22.6% | 25.1% | 26.0% | 24.9% | 25.9% | 26.4% | 26.6% | 25.6% | 25.0% | 24.8% | 23.1% | 22.7% | 20.8% | 16.5% | 14.8% | 10.5% |
| aAABb   | 0.0% | 1.1% | 1.6% | 1.9% | 2.9% | 1.8% | 2.8% | 2.8% | 2.9% | 2.9% | 2.9% | 2.9% | 2.9% | 2.9% | 2.8% | 2.8% | 2.7% | 2.7% | 2.5% | 2.2% | 1.9% | 1.9% |

#### Instances of genotypes

| Alleles | Time | 0 | 5000 | 10000 | 15000 | 20000 | 25000 | 30000 | 35000 | 40000 | 45000 | 50000 | 55000 | 60000 | 65000 | 70000 | 75000 | 80000 | 85000 | 90000 | 95000 | 100000 |
|---------|------|---|------|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| aabb    | 1.00% | 5.90 | 1.30 | 4.70 | 4.95 | 4.50 | 4.70 | 4.70 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 |
| aAabb   | 0.00 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 |
| aAABb   | 0.00 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 | 6.30 |

Case: PhenotypeLoci=0x0000000000000003; AssortativeLoci=0x0000000000000000;

Figure C.29: DM04-002 Population data
Figure C.30: DM04-002(000) Example genotype maps at 5k time-steps
Analysis of populations by genotype

Source: DM04_rareID003.csv
Max pop: 2500
Runs: 20

Total populations of genotypes for all runs

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Time steps</th>
<th>0</th>
<th>500</th>
<th>1000</th>
<th>1500</th>
<th>2000</th>
<th>2500</th>
<th>3000</th>
<th>3500</th>
<th>4000</th>
<th>4500</th>
<th>5000</th>
<th>5500</th>
<th>6000</th>
<th>6500</th>
<th>7000</th>
<th>7500</th>
<th>8000</th>
<th>8500</th>
<th>9000</th>
<th>9500</th>
<th>10000</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAbb</td>
<td></td>
<td>0</td>
<td>5000</td>
<td>10000</td>
<td>15000</td>
<td>20000</td>
<td>25000</td>
<td>30000</td>
<td>35000</td>
<td>40000</td>
<td>45000</td>
<td>50000</td>
<td>55000</td>
<td>60000</td>
<td>65000</td>
<td>70000</td>
<td>75000</td>
<td>80000</td>
<td>85000</td>
<td>90000</td>
<td>95000</td>
<td>100000</td>
</tr>
<tr>
<td>aabb</td>
<td></td>
<td>0</td>
<td>5000</td>
<td>10000</td>
<td>15000</td>
<td>20000</td>
<td>25000</td>
<td>30000</td>
<td>35000</td>
<td>40000</td>
<td>45000</td>
<td>50000</td>
<td>55000</td>
<td>60000</td>
<td>65000</td>
<td>70000</td>
<td>75000</td>
<td>80000</td>
<td>85000</td>
<td>90000</td>
<td>95000</td>
<td>100000</td>
</tr>
<tr>
<td>AAbb</td>
<td></td>
<td>0</td>
<td>5000</td>
<td>10000</td>
<td>15000</td>
<td>20000</td>
<td>25000</td>
<td>30000</td>
<td>35000</td>
<td>40000</td>
<td>45000</td>
<td>50000</td>
<td>55000</td>
<td>60000</td>
<td>65000</td>
<td>70000</td>
<td>75000</td>
<td>80000</td>
<td>85000</td>
<td>90000</td>
<td>95000</td>
<td>100000</td>
</tr>
<tr>
<td>aABb</td>
<td></td>
<td>0</td>
<td>5000</td>
<td>10000</td>
<td>15000</td>
<td>20000</td>
<td>25000</td>
<td>30000</td>
<td>35000</td>
<td>40000</td>
<td>45000</td>
<td>50000</td>
<td>55000</td>
<td>60000</td>
<td>65000</td>
<td>70000</td>
<td>75000</td>
<td>80000</td>
<td>85000</td>
<td>90000</td>
<td>95000</td>
<td>100000</td>
</tr>
</tbody>
</table>

Average for all reruns of percentage of maximum population for each genotype

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Time steps</th>
<th>0</th>
<th>500</th>
<th>1000</th>
<th>1500</th>
<th>2000</th>
<th>2500</th>
<th>3000</th>
<th>3500</th>
<th>4000</th>
<th>4500</th>
<th>5000</th>
<th>5500</th>
<th>6000</th>
<th>6500</th>
<th>7000</th>
<th>7500</th>
<th>8000</th>
<th>8500</th>
<th>9000</th>
<th>9500</th>
<th>10000</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAbb</td>
<td></td>
<td>0</td>
<td>5000</td>
<td>10000</td>
<td>15000</td>
<td>20000</td>
<td>25000</td>
<td>30000</td>
<td>35000</td>
<td>40000</td>
<td>45000</td>
<td>50000</td>
<td>55000</td>
<td>60000</td>
<td>65000</td>
<td>70000</td>
<td>75000</td>
<td>80000</td>
<td>85000</td>
<td>90000</td>
<td>95000</td>
<td>100000</td>
</tr>
<tr>
<td>aabb</td>
<td></td>
<td>0</td>
<td>5000</td>
<td>10000</td>
<td>15000</td>
<td>20000</td>
<td>25000</td>
<td>30000</td>
<td>35000</td>
<td>40000</td>
<td>45000</td>
<td>50000</td>
<td>55000</td>
<td>60000</td>
<td>65000</td>
<td>70000</td>
<td>75000</td>
<td>80000</td>
<td>85000</td>
<td>90000</td>
<td>95000</td>
<td>100000</td>
</tr>
<tr>
<td>AAbb</td>
<td></td>
<td>0</td>
<td>5000</td>
<td>10000</td>
<td>15000</td>
<td>20000</td>
<td>25000</td>
<td>30000</td>
<td>35000</td>
<td>40000</td>
<td>45000</td>
<td>50000</td>
<td>55000</td>
<td>60000</td>
<td>65000</td>
<td>70000</td>
<td>75000</td>
<td>80000</td>
<td>85000</td>
<td>90000</td>
<td>95000</td>
<td>100000</td>
</tr>
<tr>
<td>aABb</td>
<td></td>
<td>0</td>
<td>5000</td>
<td>10000</td>
<td>15000</td>
<td>20000</td>
<td>25000</td>
<td>30000</td>
<td>35000</td>
<td>40000</td>
<td>45000</td>
<td>50000</td>
<td>55000</td>
<td>60000</td>
<td>65000</td>
<td>70000</td>
<td>75000</td>
<td>80000</td>
<td>85000</td>
<td>90000</td>
<td>95000</td>
<td>100000</td>
</tr>
</tbody>
</table>

Instances of genotypes

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Time steps</th>
<th>0</th>
<th>500</th>
<th>1000</th>
<th>1500</th>
<th>2000</th>
<th>2500</th>
<th>3000</th>
<th>3500</th>
<th>4000</th>
<th>4500</th>
<th>5000</th>
<th>5500</th>
<th>6000</th>
<th>6500</th>
<th>7000</th>
<th>7500</th>
<th>8000</th>
<th>8500</th>
<th>9000</th>
<th>9500</th>
<th>10000</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAbb</td>
<td></td>
<td>0</td>
<td>5000</td>
<td>10000</td>
<td>15000</td>
<td>20000</td>
<td>25000</td>
<td>30000</td>
<td>35000</td>
<td>40000</td>
<td>45000</td>
<td>50000</td>
<td>55000</td>
<td>60000</td>
<td>65000</td>
<td>70000</td>
<td>75000</td>
<td>80000</td>
<td>85000</td>
<td>90000</td>
<td>95000</td>
<td>100000</td>
</tr>
<tr>
<td>aabb</td>
<td></td>
<td>0</td>
<td>5000</td>
<td>10000</td>
<td>15000</td>
<td>20000</td>
<td>25000</td>
<td>30000</td>
<td>35000</td>
<td>40000</td>
<td>45000</td>
<td>50000</td>
<td>55000</td>
<td>60000</td>
<td>65000</td>
<td>70000</td>
<td>75000</td>
<td>80000</td>
<td>85000</td>
<td>90000</td>
<td>95000</td>
<td>100000</td>
</tr>
<tr>
<td>AAbb</td>
<td></td>
<td>0</td>
<td>5000</td>
<td>10000</td>
<td>15000</td>
<td>20000</td>
<td>25000</td>
<td>30000</td>
<td>35000</td>
<td>40000</td>
<td>45000</td>
<td>50000</td>
<td>55000</td>
<td>60000</td>
<td>65000</td>
<td>70000</td>
<td>75000</td>
<td>80000</td>
<td>85000</td>
<td>90000</td>
<td>95000</td>
<td>100000</td>
</tr>
<tr>
<td>aABb</td>
<td></td>
<td>0</td>
<td>5000</td>
<td>10000</td>
<td>15000</td>
<td>20000</td>
<td>25000</td>
<td>30000</td>
<td>35000</td>
<td>40000</td>
<td>45000</td>
<td>50000</td>
<td>55000</td>
<td>60000</td>
<td>65000</td>
<td>70000</td>
<td>75000</td>
<td>80000</td>
<td>85000</td>
<td>90000</td>
<td>95000</td>
<td>100000</td>
</tr>
</tbody>
</table>

Figure C.31: DM04-003 Population data
Figure C.32: DM04-003(000) Example genotype maps at 5k time-steps
C.4.2 Dobzhansky-Müller sensitivity simulation (DM07)

C.4.2.1 Dobzhansky-Müller sensitivity simulation (DM07): input

The range of assortative mating index \( (q_{am}) \) and mating range \( (R_m) \) used is listed in Table C.16 on page 232. The probability of crossover \( (P_R) \) was maintained at zero in all cases.

<table>
<thead>
<tr>
<th>Run ref</th>
<th>AM index ( q_{am} )</th>
<th>Mating range ( R_m )</th>
</tr>
</thead>
<tbody>
<tr>
<td>DM07-000</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>DM07-001</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>DM07-002</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>DM07-003</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>DM07-004</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>DM07-005</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>DM07-006</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>DM07-007</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>DM07-008</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>DM07-009</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>DM07-010</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>DM08-011</td>
<td>8</td>
<td>8</td>
</tr>
</tbody>
</table>

Table C.16: Dobzhansky-Müller sensitivity simulation (DM07): inputs

The input file for this simulation is shown in Table C.17 on page 233.

The initial setting of 0x0...0FC, applied to all locations, converts to ...11111100 leaving the last two loci unmutated. The remaining loci are all mutated so that no changes will occur there which might make the outcome difficult to interpret.

The fatal phenotype (0x0..03) ensures that all A.B. forms are aborted.

Assortative mating and phenotype matching are concentrated on the last two loci by the value 0x0..03 but in order to allow some evolutionary change the phenotype matching tolerance is set to 1 so that only one of the two genes has to match at those loci for mating to be successful.

All the cases are rerun 20 times with the same random sequence used for each case.

C.4.2.2 Dobzhansky-Müller sensitivity simulation (DM07): results

These results are discussed in Section 5.4.2.2. The outputs charts and colour scheme are similar to those for the DM04 simulations, see C.4.1.2.

The results are shown in Figures C.33 to C.56. In each case the first figure shows the percentage of the population with each genotype and the second shows the development of genotype groups as a map of the lattice.
//Sympatria input file
Computer: Toshiba Satellite T110-107 Laptop;
//TITLES
Purpose: Dobzhansky-Muller varying Rm Pr and AM;
//CHANGING: 4x3=36 cases
AssortativeIndex: 0, 2, 4, 8;
MateRange: 1, 4, 8;
//LATTICE
Rows: 50;
Cols: 50;
//INITIAL SETUP
InitialSet: 0xFFFFFFFFFFFFFFFC 0xFFFFFFFFFFFFFFFC;
//BASE DATA
SeedRange: 1;
MatureAge: 1;
Seeds: 1;
CrossOvers: 0.00;
Recomb: 0.00;
Mutation: 1.00;
//ENV DEATH
EnvDeathRate: 0.01;
//PHENOTYPE FATALITY (a fatal combination of mutations)
PhenotypeFatal: 0x0000000000000003;
//PHENOTYPE COMPATIBILITY
PhenotypeLoci: 0x0000000000000003;
PhenoTolerance: 1;
//ASSORTATIVE MATING
AssortativeLoci: 0x0000000000000003;
//CYCLES
Generations: 100000; //df 20
Stops: 20; //excluding the report at time 0 //df 20
ReRuns: 20; //df 1
RandomSeed: 999; //df 0 (deprecated form 'Seed')
//rptF: Nr of Family types v Time
ReportFamilies: YES;
ReportLoci: 0xFFFFFFFFFFFFFFF;
//Colour mapping
GeneImages: Yes;
CaptureReruns: 1;
RedMask: 0x0000000000000001;
GreenMask: 0x0000000000000002;
BlueMask: 0x0000000000000004;
HapCombination: SPLIT;

Table C.17: DM07 input data
### Analysis of populations by genotype

**Source:** DM07_reportF000.csv

| init. pop. | 2500 |
| init. types | 1 |
| Reruns | 20 |

**Total populations of genotypes for all reruns**

<table>
<thead>
<tr>
<th>Alleles</th>
<th>Time</th>
<th>Aabb</th>
<th>aABb</th>
<th>aABb</th>
<th>aBBb</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5000</td>
<td>26497</td>
<td>22298</td>
<td>21131</td>
<td>19512</td>
</tr>
<tr>
<td>0</td>
<td>10000</td>
<td>26497</td>
<td>22298</td>
<td>21131</td>
<td>19512</td>
</tr>
<tr>
<td>0</td>
<td>15000</td>
<td>26497</td>
<td>22298</td>
<td>21131</td>
<td>19512</td>
</tr>
<tr>
<td>0</td>
<td>20000</td>
<td>26497</td>
<td>22298</td>
<td>21131</td>
<td>19512</td>
</tr>
<tr>
<td>0</td>
<td>25000</td>
<td>26497</td>
<td>22298</td>
<td>21131</td>
<td>19512</td>
</tr>
</tbody>
</table>

**Average for all reruns of percentage of maximum population for each genotype**

<table>
<thead>
<tr>
<th>Alleles</th>
<th>Time</th>
<th>Aabb</th>
<th>aABb</th>
<th>aABb</th>
<th>aBBb</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5000</td>
<td>100.0%</td>
<td>53.0%</td>
<td>45.4%</td>
<td>42.3%</td>
</tr>
<tr>
<td>0</td>
<td>10000</td>
<td>100.0%</td>
<td>53.0%</td>
<td>45.4%</td>
<td>42.3%</td>
</tr>
<tr>
<td>0</td>
<td>15000</td>
<td>100.0%</td>
<td>53.0%</td>
<td>45.4%</td>
<td>42.3%</td>
</tr>
<tr>
<td>0</td>
<td>20000</td>
<td>100.0%</td>
<td>53.0%</td>
<td>45.4%</td>
<td>42.3%</td>
</tr>
<tr>
<td>0</td>
<td>25000</td>
<td>100.0%</td>
<td>53.0%</td>
<td>45.4%</td>
<td>42.3%</td>
</tr>
</tbody>
</table>

**Instances of genotypes**

<table>
<thead>
<tr>
<th>Instance</th>
<th>Time</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5000</td>
<td>1.00</td>
</tr>
<tr>
<td>0</td>
<td>10000</td>
<td>1.00</td>
</tr>
<tr>
<td>0</td>
<td>15000</td>
<td>1.00</td>
</tr>
<tr>
<td>0</td>
<td>20000</td>
<td>1.00</td>
</tr>
<tr>
<td>0</td>
<td>25000</td>
<td>1.00</td>
</tr>
</tbody>
</table>

**Case:** AssortativeIndex=0; MateRange=1;
Figure C.34: DM07-000(000) Example genotype maps at 5k time-steps
## Analysis of populations by genotype

Source: DM07_report001.cav

| Instances of genotypes | Time | Avg. No. of genotypes | Case: AssortativeIndex=0; MateRange=4; | 236 |
Figure C.36: DM07-001(000) Example genotype maps at 5k time-steps
### Analysis of populations by genotype

Source: DM07_reportFig02.csv

<table>
<thead>
<tr>
<th>init. pop.</th>
<th>2500</th>
</tr>
</thead>
<tbody>
<tr>
<td>init. types</td>
<td>1</td>
</tr>
<tr>
<td>Repetitions</td>
<td>20</td>
</tr>
</tbody>
</table>

#### Total populations of genotypes for all runs

<table>
<thead>
<tr>
<th>Alleles</th>
<th>Time</th>
<th>AABB</th>
<th>AaBB</th>
<th>AaBb</th>
<th>aabb</th>
</tr>
</thead>
<tbody>
<tr>
<td>aabb</td>
<td>50000</td>
<td>32347</td>
<td>31621</td>
<td>30188</td>
<td>27705</td>
</tr>
<tr>
<td>aABb</td>
<td>0</td>
<td>6906</td>
<td>7164</td>
<td>7224</td>
<td>6562</td>
</tr>
<tr>
<td>aABb</td>
<td>0</td>
<td>6906</td>
<td>7164</td>
<td>7224</td>
<td>6562</td>
</tr>
</tbody>
</table>

#### Average for all runs of percentage of maximum population for each genotype

<table>
<thead>
<tr>
<th>Alleles</th>
<th>Time</th>
<th>aabb</th>
<th>aABb</th>
<th>AABb</th>
<th>AAbB</th>
</tr>
</thead>
<tbody>
<tr>
<td>aabb</td>
<td>100.0%</td>
<td>65.9%</td>
<td>63.2%</td>
<td>56.7%</td>
<td>55.4%</td>
</tr>
<tr>
<td>aABb</td>
<td>0.0%</td>
<td>13.8%</td>
<td>14.4%</td>
<td>13.1%</td>
<td>12.3%</td>
</tr>
<tr>
<td>AABb</td>
<td>0.0%</td>
<td>1.1%</td>
<td>1.8%</td>
<td>1.5%</td>
<td>1.9%</td>
</tr>
<tr>
<td>AAbB</td>
<td>0.0%</td>
<td>1.5%</td>
<td>2.3%</td>
<td>2.9%</td>
<td>4.4%</td>
</tr>
</tbody>
</table>

#### Instances of genotypes

<table>
<thead>
<tr>
<th>Alleles</th>
<th>Time</th>
<th>Instances</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>aAABb</td>
<td>5000</td>
<td>8.80</td>
<td>9.00</td>
</tr>
<tr>
<td>aABb</td>
<td>5000</td>
<td>8.80</td>
<td>9.00</td>
</tr>
<tr>
<td>AABb</td>
<td>5000</td>
<td>8.80</td>
<td>9.00</td>
</tr>
</tbody>
</table>

### Figure C.37: DM07-002 Population data
Figure C.38: DM07-002(000) Example genotype maps at 5k time-steps
### Analysis of populations by genotype

Source: DM07_reportF003.csv

| Init. pop. | 2500 |
| Init. types | 1 |
| Reruns | 20 |

#### Total populations of genotypes for all reruns

<table>
<thead>
<tr>
<th>Alleles</th>
<th>Time</th>
<th>Retype</th>
<th>Init. pop.</th>
<th>DM07_reportF003.csv</th>
<th>Average for all reruns of percentage of maximum population for each genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAbb</td>
<td>0.00</td>
<td>0.00</td>
<td>2500</td>
<td>100.0%</td>
<td></td>
</tr>
<tr>
<td>AaBB</td>
<td>0.00</td>
<td>0.00</td>
<td>2500</td>
<td>100.0%</td>
<td></td>
</tr>
<tr>
<td>AaBb</td>
<td>0.00</td>
<td>0.00</td>
<td>2500</td>
<td>100.0%</td>
<td></td>
</tr>
<tr>
<td>AAbb</td>
<td>0.00</td>
<td>0.00</td>
<td>2500</td>
<td>100.0%</td>
<td></td>
</tr>
<tr>
<td>AAAB</td>
<td>0.00</td>
<td>0.00</td>
<td>2500</td>
<td>100.0%</td>
<td></td>
</tr>
<tr>
<td>AAAA</td>
<td>0.00</td>
<td>0.00</td>
<td>2500</td>
<td>100.0%</td>
<td></td>
</tr>
</tbody>
</table>

#### Average for all reruns of percentage of maximum population for each genotype

<table>
<thead>
<tr>
<th>Alleles</th>
<th>Time</th>
<th>Retype</th>
<th>Init. pop.</th>
<th>DM07_reportF003.csv</th>
<th>Average for all reruns of percentage of maximum population for each genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>aabb</td>
<td>100.0%</td>
<td>72.1%</td>
<td>35.9%</td>
<td>17.2%</td>
<td></td>
</tr>
<tr>
<td>AaBB</td>
<td>0.00</td>
<td>8.5%</td>
<td>12.1%</td>
<td>5.8%</td>
<td></td>
</tr>
<tr>
<td>AaBb</td>
<td>0.00</td>
<td>4.6%</td>
<td>10.2%</td>
<td>2.1%</td>
<td></td>
</tr>
<tr>
<td>AAbb</td>
<td>0.00</td>
<td>8.8%</td>
<td>9.0%</td>
<td>5.5%</td>
<td></td>
</tr>
<tr>
<td>AABb</td>
<td>0.00</td>
<td>5.2%</td>
<td>12.3%</td>
<td>5.1%</td>
<td></td>
</tr>
<tr>
<td>AAAb</td>
<td>0.00</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td></td>
</tr>
</tbody>
</table>

#### Instances of genotypes

<table>
<thead>
<tr>
<th>Instances of genotypes</th>
<th>Time</th>
<th>Retype</th>
<th>Init. pop.</th>
<th>DM07_reportF003.csv</th>
<th>Average No of genotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAbb</td>
<td>0.00</td>
<td>0.00</td>
<td>2500</td>
<td>100.0%</td>
<td></td>
</tr>
<tr>
<td>AaBB</td>
<td>0.00</td>
<td>0.00</td>
<td>2500</td>
<td>100.0%</td>
<td></td>
</tr>
<tr>
<td>AaBb</td>
<td>0.00</td>
<td>0.00</td>
<td>2500</td>
<td>100.0%</td>
<td></td>
</tr>
<tr>
<td>AAbb</td>
<td>0.00</td>
<td>0.00</td>
<td>2500</td>
<td>100.0%</td>
<td></td>
</tr>
<tr>
<td>AAAb</td>
<td>0.00</td>
<td>0.00</td>
<td>2500</td>
<td>100.0%</td>
<td></td>
</tr>
</tbody>
</table>

Case: AssortativeIndex=2; MateRange=1;
Figure C.40: DM07-003(000) Example genotype maps at 5k time-steps
Analysis of populations by genotype

Source:DM07_reportF004.csv

| Init. pop. | 2500 |
| Init. types | 1 |
| Reruns | 20 |

Total populations of genotypes for all reruns

Reruns:
10000
20000
30000
40000
50000

AABB
AABb
AAAB
AAbb
aaBB
AaBB
aaBb
aabb

Average for all reruns of percentage of maximum population for each genotype

Time:
0
100
200
300
400
500
600
700
800
900
1000

Case: AssortativeIndex=2; MateRange=4;

Figure C.41: DM07-004 Population data
Figure C.42: DM07-004(000) Example genotype maps at 5k time-steps
Analysis of populations by genotype

Source: DM07_reportF005.csv

Init. pop. 2500
Init. types 1
Runs 20

Total populations of genotypes for all runs

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Init. pop.</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AABB</td>
<td>44851</td>
<td>0.0%</td>
</tr>
<tr>
<td>AaBb</td>
<td>44851</td>
<td>0.0%</td>
</tr>
<tr>
<td>Aabb</td>
<td>44851</td>
<td>0.0%</td>
</tr>
<tr>
<td>aaBB</td>
<td>44851</td>
<td>0.0%</td>
</tr>
<tr>
<td>AaBB</td>
<td>44851</td>
<td>0.0%</td>
</tr>
<tr>
<td>AaBb</td>
<td>44851</td>
<td>0.0%</td>
</tr>
<tr>
<td>AAAbb</td>
<td>44851</td>
<td>0.0%</td>
</tr>
<tr>
<td>AAABb</td>
<td>44851</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

Average for all runs of percentage of maximum population for each genotype

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AABB</td>
<td>100.0%</td>
</tr>
<tr>
<td>AABb</td>
<td>100.0%</td>
</tr>
<tr>
<td>AaBb</td>
<td>100.0%</td>
</tr>
<tr>
<td>AAAbb</td>
<td>100.0%</td>
</tr>
<tr>
<td>AAABb</td>
<td>100.0%</td>
</tr>
<tr>
<td>AABb</td>
<td>100.0%</td>
</tr>
<tr>
<td>AaBB</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Case: AssortativeIndex=2; MateRange=8;

Figure C.43: DM07-005 Population data
Figure C.44: DM07-005(000) Example genotype maps at 5k time-steps
Figure C.45: DM07-006 Population data
Figure C.46: DM07-006(000) Example genotype maps at 5k time-steps
## Analysis of populations by genotype

### Source: DM07_reportF007.csv

| Init. pop. | 2500 |
| Init. types | 1 |
| Recurrs | 20 |

### Total populations of genotypes for all runs

<table>
<thead>
<tr>
<th>Alleles</th>
<th>Time</th>
<th>Incidences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aabb</td>
<td>5000</td>
<td>41157 37624 32315 26331 21399 17259 13520 9181 661 4552 3463 1951 1151 349 6</td>
</tr>
<tr>
<td>aABb</td>
<td>0</td>
<td>2783 3032 3212 3302 2096 2457 2013 2128 1692 959 761 630 573 234 45</td>
</tr>
<tr>
<td>AABB</td>
<td>0</td>
<td>1508 2463 2938 5691 7784 2615 10425 12013 16362 16095 18040 38324 19290 19390 27769 29099 20043 19843</td>
</tr>
<tr>
<td>aAbb</td>
<td>0</td>
<td>2577 3215 3893 3988 3413 3293 2426 1894 1379 837 660 447 295 153 52</td>
</tr>
<tr>
<td>AAbb</td>
<td>0</td>
<td>1503 3181 3751 9911 13733 17280 21122 24280 29570 28067 28382 30497 29944 29500 30173 30189 29729 29419 29483 29541</td>
</tr>
<tr>
<td>AABb</td>
<td>0</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>aABB</td>
<td>0</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>AABB</td>
<td>0</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
</tbody>
</table>

### Average for all runs of percentage of maximum population for each genotype

<table>
<thead>
<tr>
<th>Time</th>
<th>Aabb</th>
<th>aABb</th>
<th>aABB</th>
<th>AabB</th>
<th>AabB</th>
<th>AABb</th>
<th>aAAbb</th>
<th>AAbb</th>
<th>AABb</th>
<th>AAbb</th>
<th>AABb</th>
<th>AABb</th>
<th>AABb</th>
</tr>
</thead>
<tbody>
<tr>
<td>5000</td>
<td>49288</td>
<td>48936</td>
<td>48915</td>
<td>49023</td>
<td>49536</td>
<td>48829</td>
<td>49546</td>
<td>48654</td>
<td>49560</td>
<td>48174</td>
<td>48350</td>
<td>48610</td>
<td>48896</td>
</tr>
</tbody>
</table>

### % of maximum population

<table>
<thead>
<tr>
<th>Time</th>
<th>AAbB</th>
<th>aAAbb</th>
<th>AABb</th>
<th>aABB</th>
<th>AAbb</th>
<th>AABb</th>
<th>AAbb</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

### Instances of genotypes

<table>
<thead>
<tr>
<th>Time</th>
<th>Instances</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>5000</td>
<td>2000</td>
<td>5000</td>
</tr>
</tbody>
</table>

Case: AssortativeIndex=4; MateRange=4;

**Figure C.47: DM07-007 Population data**
Figure C.48: DM07-007(000) Example genotype maps at 5k time-steps
Analysis of populations by genotype

<table>
<thead>
<tr>
<th>Allele</th>
<th>Aabb</th>
<th>AAbb</th>
<th>aaBB</th>
<th>AaBb</th>
<th>aabb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Init. pop.</td>
<td>50000</td>
<td>4357</td>
<td>4300</td>
<td>4280</td>
<td>4254</td>
</tr>
<tr>
<td>Instances</td>
<td>50000</td>
<td>2462</td>
<td>2410</td>
<td>2363</td>
<td>2314</td>
</tr>
<tr>
<td>Time steps</td>
<td>50000</td>
<td>785</td>
<td>762</td>
<td>639</td>
<td>615</td>
</tr>
<tr>
<td></td>
<td>50000</td>
<td>2325</td>
<td>2314</td>
<td>2305</td>
<td>2297</td>
</tr>
<tr>
<td></td>
<td>49526</td>
<td>614</td>
<td>844</td>
<td>1129</td>
<td>1520</td>
</tr>
</tbody>
</table>

Total populations of genotypes for all runs

<table>
<thead>
<tr>
<th>Allele</th>
<th>Aabb</th>
<th>AAbb</th>
<th>aaBB</th>
<th>AaBb</th>
<th>aabb</th>
</tr>
</thead>
<tbody>
<tr>
<td>init. pop.</td>
<td>2500</td>
<td>1</td>
<td>20</td>
<td>10000</td>
<td></td>
</tr>
<tr>
<td>Total populations</td>
<td>100.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

Average for all runs of percentage of maximum population for each genotype

<table>
<thead>
<tr>
<th>Allele</th>
<th>Aabb</th>
<th>AAbb</th>
<th>aaBB</th>
<th>AaBb</th>
<th>aabb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time steps</td>
<td>50000</td>
<td>100.0%</td>
<td>89.6%</td>
<td>85.1%</td>
<td>81.7%</td>
</tr>
<tr>
<td></td>
<td>50000</td>
<td>100.0%</td>
<td>82.0%</td>
<td>81.8%</td>
<td>81.8%</td>
</tr>
<tr>
<td></td>
<td>50000</td>
<td>100.0%</td>
<td>81.3%</td>
<td>81.2%</td>
<td>81.0%</td>
</tr>
<tr>
<td></td>
<td>50000</td>
<td>100.0%</td>
<td>81.0%</td>
<td>80.9%</td>
<td>80.8%</td>
</tr>
</tbody>
</table>

Instances of genotypes

<table>
<thead>
<tr>
<th>Instance</th>
<th>Aabb</th>
<th>AAbb</th>
<th>aaBB</th>
<th>AaBb</th>
<th>aabb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time steps</td>
<td>100.0%</td>
<td>50.0%</td>
<td>4.0%</td>
<td>1.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td></td>
<td>100.0%</td>
<td>50.0%</td>
<td>4.0%</td>
<td>1.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td></td>
<td>100.0%</td>
<td>50.0%</td>
<td>4.0%</td>
<td>1.0%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

Case: AssortativeIndex=4; MateRange=8;

Figure C.49: DM07-008 Population data
Figure C.50: DM07-008(000) Example genotype maps at 5k time-steps
Analysis of populations by genotype

Source: DM07_reportF09-csv

| Init. pop. | 2500 |
| Init. types | 1 |
| Reruns | 20 |

Total populations of genotypes for all reruns

<table>
<thead>
<tr>
<th>Alleles</th>
<th>AABB</th>
<th>AABb</th>
<th>AaBB</th>
<th>AaBb</th>
<th>AAbb</th>
<th>Aabb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instances</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Average</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

Time steps

Case: AssortativeIndex=8; MateRange=1;

Figure C.51: DM07-009 Population data
Figure C.52: DM07-009(000) Example genotype maps at 5k time-steps
Analysis of populations by genotype

Source: DM07_report010.csv

<table>
<thead>
<tr>
<th>Initial population</th>
<th>2500</th>
<th>Initial genotypes</th>
<th>1</th>
<th>Recurrence</th>
<th>20</th>
</tr>
</thead>
</table>

Total populations of genotypes for all runs

<table>
<thead>
<tr>
<th>Allele Combination</th>
<th>Time</th>
<th>5000</th>
<th>10000</th>
<th>15000</th>
<th>20000</th>
<th>25000</th>
<th>30000</th>
<th>35000</th>
<th>40000</th>
<th>45000</th>
<th>50000</th>
<th>55000</th>
</tr>
</thead>
<tbody>
<tr>
<td>AABB</td>
<td>0</td>
<td>5000</td>
<td>10000</td>
<td>15000</td>
<td>20000</td>
<td>25000</td>
<td>30000</td>
<td>35000</td>
<td>40000</td>
<td>45000</td>
<td>50000</td>
<td>55000</td>
</tr>
<tr>
<td>AaBb</td>
<td>0</td>
<td>5000</td>
<td>10000</td>
<td>15000</td>
<td>20000</td>
<td>25000</td>
<td>30000</td>
<td>35000</td>
<td>40000</td>
<td>45000</td>
<td>50000</td>
<td>55000</td>
</tr>
<tr>
<td>AaBB</td>
<td>0</td>
<td>5000</td>
<td>10000</td>
<td>15000</td>
<td>20000</td>
<td>25000</td>
<td>30000</td>
<td>35000</td>
<td>40000</td>
<td>45000</td>
<td>50000</td>
<td>55000</td>
</tr>
<tr>
<td>aabb</td>
<td>0</td>
<td>5000</td>
<td>10000</td>
<td>15000</td>
<td>20000</td>
<td>25000</td>
<td>30000</td>
<td>35000</td>
<td>40000</td>
<td>45000</td>
<td>50000</td>
<td>55000</td>
</tr>
</tbody>
</table>

Average for all runs of percentage of maximum population for each genotype

<table>
<thead>
<tr>
<th>Allele Combination</th>
<th>Time</th>
<th>0</th>
<th>5000</th>
<th>10000</th>
<th>15000</th>
<th>20000</th>
<th>25000</th>
<th>30000</th>
<th>35000</th>
<th>40000</th>
<th>45000</th>
<th>50000</th>
<th>55000</th>
</tr>
</thead>
<tbody>
<tr>
<td>AABB</td>
<td>0</td>
<td>100.0%</td>
<td>51.2%</td>
<td>27.9%</td>
<td>15.7%</td>
<td>7.2%</td>
<td>3.9%</td>
<td>2.1%</td>
<td>1.1%</td>
<td>0.8%</td>
<td>0.2%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>AaBb</td>
<td>0</td>
<td>0.0%</td>
<td>5.2%</td>
<td>4.5%</td>
<td>3.0%</td>
<td>2.1%</td>
<td>1.9%</td>
<td>0.5%</td>
<td>0.3%</td>
<td>0.2%</td>
<td>0.1%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>AaBB</td>
<td>0</td>
<td>0.0%</td>
<td>18.1%</td>
<td>29.8%</td>
<td>31.1%</td>
<td>41.5%</td>
<td>44.8%</td>
<td>46.9%</td>
<td>47.6%</td>
<td>47.9%</td>
<td>48.4%</td>
<td>48.9%</td>
<td>49.7%</td>
</tr>
<tr>
<td>aabb</td>
<td>0</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>aABb</td>
<td>0</td>
<td>0.0%</td>
<td>19.6%</td>
<td>32.3%</td>
<td>40.2%</td>
<td>44.6%</td>
<td>46.8%</td>
<td>48.6%</td>
<td>49.7%</td>
<td>50.2%</td>
<td>50.1%</td>
<td>49.3%</td>
<td>49.3%</td>
</tr>
<tr>
<td>aABB</td>
<td>0</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>AAAB</td>
<td>0</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>AAbB</td>
<td>0</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>AAAB</td>
<td>0</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>AAABB</td>
<td>0</td>
<td>100.0%</td>
<td>99.0%</td>
<td>99.0%</td>
<td>99.0%</td>
<td>99.0%</td>
<td>99.0%</td>
<td>99.0%</td>
<td>99.0%</td>
<td>99.0%</td>
<td>99.0%</td>
<td>99.0%</td>
<td>99.0%</td>
</tr>
</tbody>
</table>

Instances of genotypes

<table>
<thead>
<tr>
<th>Allele Combination</th>
<th>Time</th>
<th>0</th>
<th>5000</th>
<th>10000</th>
<th>15000</th>
<th>20000</th>
<th>25000</th>
<th>30000</th>
<th>35000</th>
<th>40000</th>
<th>45000</th>
<th>50000</th>
<th>55000</th>
</tr>
</thead>
<tbody>
<tr>
<td>AABB</td>
<td>0</td>
<td>100.0%</td>
<td>99.0%</td>
<td>99.0%</td>
<td>99.0%</td>
<td>99.0%</td>
<td>99.0%</td>
<td>99.0%</td>
<td>99.0%</td>
<td>99.0%</td>
<td>99.0%</td>
<td>99.0%</td>
<td>99.0%</td>
</tr>
<tr>
<td>AaBb</td>
<td>0</td>
<td>1587</td>
<td>25577</td>
<td>36547</td>
<td>45014</td>
<td>54481</td>
<td>63948</td>
<td>73416</td>
<td>82883</td>
<td>92350</td>
<td>101817</td>
<td>111284</td>
<td>120751</td>
</tr>
<tr>
<td>AaBB</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>aabb</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>aABb</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>aABB</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AAAB</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AAbB</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AAAB</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AAABB</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AAABB</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Case: AssortativeIndex=8; MateRange=4;
Figure C.54: DM07-010(000) Example genotype maps at 5k time-steps
### Analysis of populations by genotype

**Source:** DM07_reportF011.csv

#### Instance pop:
- **init. pop.:** 2500
- **init. types:** 1
- **Reruns:** 20

#### Total populations of genotypes for all reruns:

<table>
<thead>
<tr>
<th>Genotypes</th>
<th>Init. pop.</th>
<th>DM07_reportF011.csv</th>
<th>Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>AABB</td>
<td>50000</td>
<td>25000</td>
<td>0</td>
</tr>
<tr>
<td>AaBB</td>
<td>0</td>
<td>2796</td>
<td>0</td>
</tr>
<tr>
<td>Aabb</td>
<td>0</td>
<td>8773</td>
<td>0</td>
</tr>
<tr>
<td>AaBb</td>
<td>0</td>
<td>3051</td>
<td>0</td>
</tr>
<tr>
<td>aaBB</td>
<td>0</td>
<td>9964</td>
<td>0</td>
</tr>
<tr>
<td>aABb</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>aAABb</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AAABB</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>aAABB</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AABB</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AaBB</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AAABb</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AAaBB</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AaABb</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AABb</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AAaBb</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AaABb</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>aABb</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>aAABb</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AAaBb</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AaABb</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AAaBb</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AaABb</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AAaBb</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AaABb</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AAaBb</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AaABb</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AAaBb</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AaABb</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AAaBb</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AaABb</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

#### Average for all reruns of percentage of maximum population for each genotype:

<table>
<thead>
<tr>
<th>Genotypes</th>
<th>Time 0</th>
<th>Time 5000</th>
<th>Time 10000</th>
<th>Time 15000</th>
<th>Time 20000</th>
<th>Time 30000</th>
<th>Time 50000</th>
<th>Time 100000</th>
</tr>
</thead>
<tbody>
<tr>
<td>AABB</td>
<td>100.0%</td>
<td>50.0%</td>
<td>25.0%</td>
<td>12.5%</td>
<td>6.3%</td>
<td>3.2%</td>
<td>1.6%</td>
<td>0.8%</td>
</tr>
<tr>
<td>Aabb</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>AAaBb</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>aAABB</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>AAAbb</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

#### Instances of genotypes:

<table>
<thead>
<tr>
<th>Instances</th>
<th>Time 0</th>
<th>Time 5000</th>
<th>Time 10000</th>
<th>Time 15000</th>
<th>Time 20000</th>
<th>Time 30000</th>
<th>Time 50000</th>
<th>Time 100000</th>
</tr>
</thead>
<tbody>
<tr>
<td>AABB</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>AaBB</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Aabb</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>AaBb</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

---

**Figure C.55:** DM07-011 Population data
Figure C.56: DM07-011(000) Example genotype maps at 5k time-steps
C.4.3 Dobzhansky-Müller de-linking simulation (DM08)

This simulation to test the effect of the relative separation on the chromosome of the key loci in the Dobzhansky-Müller model. The simulation is discussed in Section 5.4.3 of the main thesis.

C.4.3.1 Dobzhansky-Müller de-linking simulation (DM08): input

This set of runs tested for any influence of crossover on the outcome. The principal values used are summarized in Table C.18 on page 258.

<table>
<thead>
<tr>
<th>Run ref</th>
<th>$P_R$</th>
<th>$q_{am}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DM08-000</td>
<td>0.20</td>
<td>4</td>
</tr>
<tr>
<td>DM08-001</td>
<td>0.20</td>
<td>8</td>
</tr>
<tr>
<td>DM08-002</td>
<td>0.40</td>
<td>4</td>
</tr>
<tr>
<td>DM08-003</td>
<td>0.40</td>
<td>8</td>
</tr>
<tr>
<td>DM08-004</td>
<td>0.80</td>
<td>4</td>
</tr>
<tr>
<td>DM08-005</td>
<td>0.80</td>
<td>8</td>
</tr>
</tbody>
</table>

Table C.18: DM08 Assumptions for testing separated loci for Aa and Bb

The input file for this simulation is shown in Table C.19 on page 259.

C.4.3.2 Dobzhansky-Müller de-linking simulation (DM08): results

The results are shown in Figures Figure C.57 on page 260 to Figure C.68 on page 271. In each case the first figure shows the percentage of the population with each genotype and the second shows the development of genotype groups as a map of the lattice.
//Sympatria input file
Computer: Toshiba Satellite T110-107 Laptop;

//TITLES
Purpose: Dobzhansky-Muller with separated loci;

//LATTICE
Rows: 50;
Cols: 50;

//INITIAL SETUP
InitialSet: 0xFFFF7FFFFFFEFFFF 0xFFFF7FFFFFFEFFFF;

//BASE DATA
MateRange: 1; //for rapid response
SeedRange: 1;
MatureAge: 1;
Seeds: 1;
Recomb: 0.20, 0.40, 0.80;
CrossOvers: 1.00;
Mutation: 1.00;

//PHENOTYPE FATALITY (a fatal combination of mutations)
PhenotypeFatal: 0x0000800000010000;

//PHENOTYPE COMPATIBILITY
PhenotypeLoci: 0x0000800000010000;
PhenoTolerance: 1;

//ASSORTATIVE MATING
AssortativeLoci: 0x0000800000010000;
AssortativeIndex: 4, 8;

//CYCLES
Generations: 100000; //df 20
Stops: 20; //excluding the report at time 0 //df 20
ReRuns: 20; //df 1
RandomSeed: 999; //df 0 (deprecated form 'Seed')

//rptF: Nr of Family types v Time
ReportFamilies: YES;
ReportLoci: 0xFFFFFFFFFFFFFFFF;

//Colour mapping
GeneImages: Yes;
CaptureReruns: 1;
RedMask: 0x0000000000010000;
GreenMask: 0x0000000000000000;
BlueMask: 0x0000000000000000;
HapCombination: SPLIT;

Table C.19: DM08 input data
Analysis of populations by genotype

Source: DM08_reportF000.csv

Max pop. 2500
Reruns 20

Total populations of genotypes for all reruns

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Instances</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAbb</td>
<td>0</td>
<td>5000</td>
</tr>
<tr>
<td>aaBb</td>
<td>0</td>
<td>5000</td>
</tr>
<tr>
<td>Aabb</td>
<td>0</td>
<td>5000</td>
</tr>
<tr>
<td>aabb</td>
<td>0</td>
<td>5000</td>
</tr>
<tr>
<td>AABB</td>
<td>0</td>
<td>5000</td>
</tr>
<tr>
<td>aABb</td>
<td>0</td>
<td>5000</td>
</tr>
<tr>
<td>aAbb</td>
<td>0</td>
<td>5000</td>
</tr>
</tbody>
</table>

Average for all reruns of percentage of maximum population for each genotype

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAbb</td>
<td>100.0%</td>
</tr>
<tr>
<td>aaBb</td>
<td>10.0%</td>
</tr>
<tr>
<td>Aabb</td>
<td>20.0%</td>
</tr>
<tr>
<td>aabb</td>
<td>30.0%</td>
</tr>
<tr>
<td>AABB</td>
<td>40.0%</td>
</tr>
<tr>
<td>aABb</td>
<td>50.0%</td>
</tr>
<tr>
<td>aAbb</td>
<td>60.0%</td>
</tr>
<tr>
<td>aABB</td>
<td>70.0%</td>
</tr>
</tbody>
</table>

Instances of genotypes

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Instances</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>aabb</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>AAbb</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>AAAbb</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>aABb</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Case: Recomb=0.20; AssortativeIndex=4;

Figure C.57: DM08-000 Population data
Figure C.58: DM08-000(000) Example genotype maps at 5k time-steps
Analysis of populations by genotype

Source: DM08_reportF001.csv

Max pop: 2500
Reps: 20

Total populations of genotypes for all runs

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Time</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAbb</td>
<td>5000</td>
<td>355620</td>
</tr>
<tr>
<td>aabb</td>
<td>5000</td>
<td>35467</td>
</tr>
<tr>
<td>AaBb</td>
<td>5000</td>
<td>35362</td>
</tr>
<tr>
<td>aaBB</td>
<td>5000</td>
<td>35267</td>
</tr>
</tbody>
</table>

Average for all runs of maximum population for each genotype

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Time</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>aabb</td>
<td>5000</td>
<td>14987</td>
</tr>
<tr>
<td>AAbb</td>
<td>5000</td>
<td>14987</td>
</tr>
<tr>
<td>aaBB</td>
<td>5000</td>
<td>14987</td>
</tr>
</tbody>
</table>

Instances of genotypes

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Time</th>
<th>Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>aabb</td>
<td>5000</td>
<td>355620</td>
</tr>
<tr>
<td>AAbb</td>
<td>5000</td>
<td>35467</td>
</tr>
<tr>
<td>AaBb</td>
<td>5000</td>
<td>35362</td>
</tr>
<tr>
<td>aaBB</td>
<td>5000</td>
<td>35267</td>
</tr>
</tbody>
</table>

Case: Recomb=0.20; AssortativeIndex=8;

Figure C.59: DM08-001 Population data
Figure C.60: DM08-001(000) Example genotype maps at 5k time-steps
Analysis of populations by genotype

Source: DM08_reportF002.csv
Max pop. 2500
Reruns 20

Total populations of genotypes for all reruns

<table>
<thead>
<tr>
<th>Time</th>
<th>0</th>
<th>5000</th>
<th>10000</th>
<th>15000</th>
<th>20000</th>
<th>30000</th>
<th>50000</th>
<th>100000</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>1000</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>10000</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Average for all reruns of percentage of maximum population for each genotype

<table>
<thead>
<tr>
<th>Time</th>
<th>0</th>
<th>5000</th>
<th>10000</th>
<th>15000</th>
<th>20000</th>
<th>30000</th>
<th>50000</th>
<th>100000</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>1000</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>10000</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Instances of genotypes

<table>
<thead>
<tr>
<th>Time</th>
<th>0</th>
<th>5000</th>
<th>10000</th>
<th>15000</th>
<th>20000</th>
<th>30000</th>
<th>50000</th>
<th>100000</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>1000</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>10000</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Case: Recomb=0.40; AssortativeIndex=4;

Figure C.61: DM08-002 Population data
Figure C.62: DM08-002(000) Example genotype maps at 5k time-steps
Analysis of populations by genotype

Source: DM08_reportF003.csv
Max pop. 2500
Reruns 20

Total populations of genotypes for all reruns

| Alleles | 0 | 5000 | 10000 | 15000 | 20000 | 25000 | 30000 | 35000 | 40000 | 45000 | 50000 | 55000 | 60000 | 65000 | 70000 | 75000 | 80000 | 85000 | 90000 | 95000 | 100000 |
|---------|---|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| aabb    | 0 | 5375 | 24694 | 15659 | 9344  | 4398  | 1684  | 476   | 343   | 71    | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
| AAbb    | 0 | 2127 | 2462  | 1972  | 1309  | 856   | 413   | 115   | 76    | 38    | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
| aabb    | 0 | 2222 | 2175  | 2460  | 1691  | 1099  | 476   | 225   | 97    | 40    | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |
| AAbb    | 0 | 4868 | 9803  | 14188 | 17172 | 18724 | 20197 | 19862 | 19451 | 20117 | 20577 | 20748 | 21054 | 20644 | 20546 | 20126 | 20128 | 20061 | 19720 | 19109 | 19343 |
| aAbb    | 0 | 6213 | 10145 | 15196 | 20023 | 24663 | 29701 | 28851 | 29568 | 29237 | 28918 | 28599 | 28458 | 28030 | 28182 | 28394 | 28439 | 29783 | 30487 | 30160 |
| 50000   | 49517 | 49500 | 49475 | 49539 | 49540 | 49471 | 49479 | 49485 | 49503 | 49495 | 49447 | 49512 | 49492 | 49488 | 49483 | 49522 | 49406 | 49513 | 49503 |

Average for all reruns of percentage of maximum population for each genotype

Time: % of imum population

Average No of genotypes

Case: Recomb=0.40; AssortativeIndex=8;

Figure C.63: DM08-003 Population data
Figure C.64: DM08-003(000) Example genotype maps at 5k time-steps
Analysis of populations by genotype

Source: DM08_reportF004.csv

Max pop.
Reruns 20

Total populations of genotypes for all reruns

<table>
<thead>
<tr>
<th>Alleles</th>
<th>Time</th>
<th>Max pop.</th>
</tr>
</thead>
<tbody>
<tr>
<td>aabb</td>
<td>0</td>
<td>50000</td>
</tr>
<tr>
<td>AAbb</td>
<td>0</td>
<td>45524</td>
</tr>
<tr>
<td>aaBb</td>
<td>0</td>
<td>45430</td>
</tr>
<tr>
<td>Aabb</td>
<td>0</td>
<td>45035</td>
</tr>
<tr>
<td>aBb</td>
<td>0</td>
<td>45010</td>
</tr>
<tr>
<td>AAb</td>
<td>0</td>
<td>45035</td>
</tr>
<tr>
<td>aB</td>
<td>0</td>
<td>50000</td>
</tr>
<tr>
<td>AB</td>
<td>0</td>
<td>50000</td>
</tr>
<tr>
<td>a</td>
<td>0</td>
<td>50000</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>50000</td>
</tr>
<tr>
<td>b</td>
<td>0</td>
<td>50000</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>50000</td>
</tr>
</tbody>
</table>

Average for all reruns of percentage of maximum population for each genotype

<table>
<thead>
<tr>
<th>Alleles</th>
<th>Time</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>aabb</td>
<td>0</td>
<td>50.2%</td>
</tr>
<tr>
<td>AAbb</td>
<td>0</td>
<td>50.2%</td>
</tr>
<tr>
<td>aaBb</td>
<td>0</td>
<td>50.2%</td>
</tr>
<tr>
<td>Aabb</td>
<td>0</td>
<td>50.2%</td>
</tr>
<tr>
<td>aBb</td>
<td>0</td>
<td>50.2%</td>
</tr>
<tr>
<td>AAb</td>
<td>0</td>
<td>50.2%</td>
</tr>
<tr>
<td>aB</td>
<td>0</td>
<td>50.2%</td>
</tr>
<tr>
<td>AB</td>
<td>0</td>
<td>50.2%</td>
</tr>
<tr>
<td>a</td>
<td>0</td>
<td>50.2%</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>50.2%</td>
</tr>
<tr>
<td>b</td>
<td>0</td>
<td>50.2%</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>50.2%</td>
</tr>
</tbody>
</table>

Instances of genotypes

<table>
<thead>
<tr>
<th>Alleles</th>
<th>Time</th>
<th>Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>aabb</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>AAbb</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>aaBb</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>Aabb</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>aBb</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>AAb</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>aB</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>AB</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>a</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>b</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>25</td>
</tr>
</tbody>
</table>

Case: Recomb=0.80; AssortativeIndex=4;

Figure C.65: DM08-004 Population data
Figure C.66: DM08-004(000) Example genotype maps at 5k time-steps
Figure C.67: DM08-005 Population data
Figure C.68: DM08-005(000) Example genotype maps at 5k time-steps
C.4.4 Dobzhansky-Müller with no fatal phenotype (DM09)

This simulation is discussed in Section 5.4.4.

C.4.4.1 Dobzhansky-Müller with no fatal phenotype (DM09): input

The input for this simulation is shown in Table C.20 on page 272.

```
//Sympatria input file
Computer: Toshiba Satellite T110-107 Laptop;
//TITLES
Purpose: Dobzhansky-Muller with no fatal phenotype;
//LATTICE
Rows: 50;
Cols: 50;
//INITIAL SETUP
InitialSet: 0xFFFF7FFFFFFEFFFF 0xFFFF7FFFFFFEFFFF;
//BASE DATA
MateRange: 1;
SeedRange: 1;
MatureAge: 1;
Seeds: 1;
Recomb: 0.20;
CrossOvers: 1.00;
Mutation: 1.00;
//PHENOTYPE FATALITY (a fatal combination of mutations)
//None
//PHENOTYPE COMPATIBILITY
PhenotypeLoci: 0x0000800000010000;
PhenoTolerance: 1;
//ASSORTATIVE MATING
AssortativeLoci: 0x0000800000010000;
AssortativeIndex: 4;
//CYCLES
Generations: 100000;
Stops: 20;
ReRuns: 1;
RandomSeed: 999;
//rptF: Nr of Family types v Time
ReportFamilies: YES;
ReportLoci: 0xFFFFFFFFFFFFFFFF;
//Colour mapping
GeneImages: Yes;
CaptureReruns: 1;
RedMask: 0x0000000000010000;
GreenMask: 0x0000800000000000;
BlueMask: 0x0000000100000000;
HapCombination: SPLIT;
```

Table C.20: DM09 input data
C.4.4.2 Dobzhansky-Müller with no fatal phenotype (DM09): results

The results are shown in Figures C.69 & C.70. The first figure shows the percentage of the population with each genotype and the second shows the development of genotype groups as a map of the lattice.

With no exclusion of the $AB$ phenotype, it can be seen that $AABB$ becomes the only genotype remaining after 50,000 time-steps.
Analysis of populations by genotype

Source: DM08_report000.cav

Init. pop. 2500
Init. types 1
Reruns 1

Total populations of genotypes for all reruns

AABB
AABb
AaBb
aaBB
aabb

Instances
Average

<table>
<thead>
<tr>
<th>N</th>
<th>Time</th>
<th>Population</th>
<th>A</th>
<th>B</th>
<th>a</th>
<th>b</th>
<th>aabb</th>
<th>AABB</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5000</td>
<td>10000</td>
<td>15000</td>
<td>20000</td>
<td>25000</td>
<td>30000</td>
<td>35000</td>
<td>40000</td>
</tr>
<tr>
<td>0</td>
<td>5000</td>
<td>10000</td>
<td>15000</td>
<td>20000</td>
<td>25000</td>
<td>30000</td>
<td>35000</td>
<td>40000</td>
</tr>
</tbody>
</table>

Average for all reruns of percentage of maximum population for each genotype

AABB
AABb
AaBb
aaBB

Instances of genotypes

AABB
AABb
AaBb
aaBB

Case:

Figure C.69: DM09-000 Population data
Figure C.70: DM09-000(000) Example genotype maps at 5k time-steps
C.4.5 Dobzhansky-Müller with no added mutations (DM10)

This simulation is discussed in Section 5.4.4.

C.4.5.1 Dobzhansky-Müller with no added mutations (DM10): input

The input for this simulation is shown in Table C.21 on page 276.

```plaintext
//Sympatria input file
Computer: Toshiba Satellite T110-107 Laptop;
//TITLES
Purpose: Dobzhansky-Muller with no mutation;
//LATTICE
Rows: 50;
Cols: 50;
//INITIAL SETUP
InitialSet: 0xFFFF7FFFFFFEFFFF 0xFFFF7FFFFFFEFFFF;
//BASE DATA
MateRange: 1; //for rapid response
SeedRange: 1;
MatureAge: 1;
Seeds: 1;
Recomb: 0.20;
CrossOvers: 1.00;
Mutation: 0.00;
//PHENOTYPE FATALITY (a fatal combination of mutations)
//None
//PHENOTYPE COMPATIBILITY
PhenotypeLoci: 0x0000800000000000;
PhenoTolerance: 1;
//ASSORTATIVE MATING
AssortativeLoci: 0x0000800000000000;
AssortativeIndex: 4;
//CYCLES
Generations: 10000;
Stops: 20;
ReRuns: 1;
RandomSeed: 999;
//rptF: Nr of Family types v Time
ReportFamilies: YES;
ReportLoci: 0xFFFFFFFFFFFFFFFF;
//Colour mapping
GeneImages: Yes;
CaptureReRuns: 1;
RedMask: 0x0008000000010000;
GreenMask: 0x0008000000000000;
BlueMask: 0x0008000000000000;
HapCombination: SPLIT;
```

Table C.21: DM10 input data
C.4.5.2 Dobzhansky-Müller with no added mutations (DM10): results

The results are shown in Figures C.71 & C.72. The first figure shows the percentage of the population with each genotype and the second shows the development of genotype groups as a map of the lattice.

As expected, with no additional mutations the population remains only genotype $aabb$. 
Analysis of populations by genotype

Source: DM10_reportF000.csv

Init. pop. 2500
Init. types 1
Reruns 1

Total populations of genotypes for all reruns

| Alleles | Time | 500 | 1000 | 1500 | 2000 | 2500 | 3000 | 3500 | 4000 | 4500 | 5000 | 5500 | 6000 | 6500 | 7000 | 7500 | 8000 | 8500 | 9000 | 9500 | 10000 |
|---------|------|-----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| aabb    | 2500 | 2459 | 2481 | 2476 | 2474 | 2478 | 2472 | 2462 | 2483 | 2476 | 2481 | 2480 | 2477 | 2478 | 2461 | 2481 | 2481 | 2476 | 2480 | 2479 |
| AaBb    | 0     | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    |
| Aabb    | 0     | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    |
| AAbb    | 0     | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    |
| AaBb    | 0     | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    |
| AABb    | 0     | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    |
| AaBb    | 0     | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    |
| AABb    | 0     | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    |
| AABB    | 0     | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    |

Figure C.71: DM10-000 Population data

278
Figure C.72: DM10-000(000) Example genotype maps at 5k time-steps
C.4.6 Dobzhansky-Müller starting from $Aabb$ and $aaBb$ (DM11)

This simulation is discussed in Section 5.4.5.

C.4.6.1 Dobzhansky-Müller starting from $Aabb$ and $aaBb$ (DM11): input

The input for this simulation is shown in Table C.22 on page 280.

```plaintext
//Sympatria input file
Computer: Toshiba Satellite T110-107 Laptop;

//TITLES
Purpose: From Aabb and aaBb with no extra mutation no AB exclusion with and
without AM;

//LATTICE
Rows: 50;
Cols: 50;

//INITIAL SETUP Aabb and aaBb
InitialSet: LeftRight FFFFFFFFEEEEEEEEEEEEEEEEEEEEEEE
EEEEEFFFFFFFFFFFFFFFFFFFFF;

//BASE DATA
MateRange: 1;
SeedRange: 1;
MatureAge: 1;
Seeds: 1;
Recomb: 0.80;
CrossOvers: 1.00;
Mutation: 0.00; //NONE

//PHENOTYPE FATALITY (a fatal combination of mutations)
//NONE

//PHENOTYPE COMPATIBILITY
//NONE

//ASSORTATIVE MATING
AssortativeLocis: 0x0000800000010000;
AssortativeIndex: 0, 4; //OFF and ON

//CYCLES
Generations: 100000;
Stops: 20;
ReRuns: 20;
RandomSeed: 999;

//rptF: Nr of Family types v Time
ReportFamilies: YES;
ReportLocis: 0xFFFFFFFFFFFFFFFF;

//Colour mapping
GeneImages: Yes;
CaptureReRuns: 1;
RedMask: 0x0000800000018000;
GreenMask: 0x0000000000010000;
BlueMask: 0x0000000000010000;
HapCombination: SPLIT;
```

Table C.22: DM11 input data
The results are shown in Figures C.73 to C.76. In each case the first figure shows the percentage of the population with each genotype and the second shows the development of genotype groups as a map of the lattice.
### Analysis of populations by genotype

**Source:** DM11_reportF000.csv

#### Init. pop.
- 2500
- Init. types: 2
- Reruns: 20

**Total populations of genotypes for all reruns**

<table>
<thead>
<tr>
<th>Genotypes</th>
<th>Time</th>
<th>Init. pop.</th>
<th>Average for all reruns</th>
</tr>
</thead>
<tbody>
<tr>
<td>AABB</td>
<td>0</td>
<td>16774</td>
<td>8855</td>
</tr>
<tr>
<td>AABb</td>
<td>0</td>
<td>6352</td>
<td>7745</td>
</tr>
<tr>
<td>AaBB</td>
<td>0</td>
<td>6348</td>
<td>7745</td>
</tr>
<tr>
<td>aaBb</td>
<td>0</td>
<td>6348</td>
<td>7746</td>
</tr>
<tr>
<td>aabb</td>
<td>0</td>
<td>6348</td>
<td>7746</td>
</tr>
</tbody>
</table>

**Average No of genotypes**

- 100.0%

---

### Case: AssortativeIndex=0

**Figure C.73: DM11-000 Population data**

282
Figure C.74: DM11-000(000) Example genotype maps at 5k time-steps
**Analysis of populations by genotype**

**Source:** DM11_reportFO01.cav

**Init. pop:** 2500  
**Init. types:** 2  
**Repeats:** 20

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Total population</th>
<th>Percentage of maximum population</th>
</tr>
</thead>
<tbody>
<tr>
<td>AABB</td>
<td>100.0%</td>
<td></td>
</tr>
<tr>
<td>AaBB</td>
<td>50.0%</td>
<td></td>
</tr>
<tr>
<td>AAbb</td>
<td>0.0%</td>
<td></td>
</tr>
<tr>
<td>aabb</td>
<td>0.0%</td>
<td></td>
</tr>
</tbody>
</table>

**Average for all runs of percentage of maximum population for each genotype**

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Time steps</th>
<th>Percentage of maximum population</th>
</tr>
</thead>
<tbody>
<tr>
<td>AABB</td>
<td>10.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>AaBB</td>
<td>20.0%</td>
<td>50.0%</td>
</tr>
<tr>
<td>AAbb</td>
<td>50.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>aabb</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

**Figure C.75: DM11-001 Population data**

Case: AssortativeIndex=4;
Figure C.76: DM11-001(000) Example genotype maps at 5k time-steps
C.4.7 Dobzhansky-Müller with initial small areas of $Aabb$ and $aaBb$ (DM12)

This simulation is discussed in Section 5.4.6

C.4.7.1 Dobzhansky-Müller with initial small areas of $Aabb$ and $aaBb$ (DM12): input

A new initial population was designed which sets the whole lattice to the aabb genotype and then sets small areas in the top left and bottom right corners to Aabb and aaBb respectively.

The input for this simulation is shown in Table C.23 on page 287. The objective was to show the development of the genotypes starting from the corners and therefore short runs of 20,000 time-steps were used.

The probability of additional mutation in each gamete was varied from 0.00 to 0.50 over the six runs and the genotype maps recorded.

C.4.7.2 Dobzhansky-Müller with initial small areas of $Aabb$ and $aaBb$ (DM12): results

The sets of genotype maps are laid out in figures C.78 to C.88 which show the impact of progressively increasing the rate of additional mutation.
//Sympatria input file
Computer: Toshiba Satellite T110-107 Laptop;

//TITLES
Purpose: aabb with varying extra mutation AB exclusion but no crossover;

//LATTICE
Rows: 50;
Cols: 50;

//INITIAL SETUP aabb with Aabb and aaBb in 5x5 corners
InitialSet: ABcorners 0x0000000000000001 0x0000000000000002 5;

//BASE DATA
MateRange: 8;
SeedRange: 1;
MatureAge: 1;
Seeds: 1;
Recomb: 0.00;
CrossOvers: 0.00;
Mutation: 0.00, 0.10, 0.20, 0.30, 0.40, 0.50;

//PHENOTYPE FATALITY (a fatal combination of mutations)
PhenotypeFatal: 0x0000000000000003;

//PHENOTYPE COMPATIBILITY
PhenotypeLoci: 0x0000000000000003;
PhenoTolerance: 1;

//ASSORTATIVE MATING
AssortativeLoci: 0x0000000000000003;
AssortativeIndex: 8;

//CYCLES
Generations: 20000;
Stops: 20;
ReRuns: 20;
RandomSeed: 999;

//rpf: Nr of Family types v Time
ReportFamilies: YES;
ReportLoci: 0xFFFFFFFFFFFFFFFF;

//Colour mapping
GeneImages: YES;
CaptureReruns: 1;
RedMask: 0x0000000000000001;
GreenMask: 0x0000000000000002;
BlueMask: 0x0000000000000004;
HapCombination: SPLIT;

Table C.23: DM12 input data
Analysis of populations by genotype

Source: DM12_report000.csv

Init. pop. 2500
Init. types 3
Reruns 20

Total populations of genotypes for all reruns

<table>
<thead>
<tr>
<th>Time</th>
<th>Aabb</th>
<th>Aabb</th>
<th>Aabb</th>
<th>Aabb</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4871</td>
<td>1491</td>
<td>4166</td>
<td>4871</td>
</tr>
<tr>
<td>1</td>
<td>1491</td>
<td>4166</td>
<td>4871</td>
<td>1491</td>
</tr>
<tr>
<td>2</td>
<td>4166</td>
<td>4871</td>
<td>1491</td>
<td>4166</td>
</tr>
<tr>
<td>3</td>
<td>4871</td>
<td>1491</td>
<td>4166</td>
<td>4871</td>
</tr>
<tr>
<td>4</td>
<td>1491</td>
<td>4166</td>
<td>4871</td>
<td>1491</td>
</tr>
<tr>
<td>5</td>
<td>4166</td>
<td>4871</td>
<td>1491</td>
<td>4166</td>
</tr>
</tbody>
</table>

Average for all reruns of percentage of maximum population for each genotype

<table>
<thead>
<tr>
<th>Time</th>
<th>Aabb</th>
<th>Aabb</th>
<th>Aabb</th>
<th>Aabb</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>98.0%</td>
<td>97.7%</td>
<td>98.3%</td>
<td>98.4%</td>
</tr>
<tr>
<td>1</td>
<td>98.0%</td>
<td>97.7%</td>
<td>98.3%</td>
<td>98.4%</td>
</tr>
<tr>
<td>2</td>
<td>98.0%</td>
<td>97.7%</td>
<td>98.3%</td>
<td>98.4%</td>
</tr>
<tr>
<td>3</td>
<td>98.0%</td>
<td>97.7%</td>
<td>98.3%</td>
<td>98.4%</td>
</tr>
<tr>
<td>4</td>
<td>98.0%</td>
<td>97.7%</td>
<td>98.3%</td>
<td>98.4%</td>
</tr>
<tr>
<td>5</td>
<td>98.0%</td>
<td>97.7%</td>
<td>98.3%</td>
<td>98.4%</td>
</tr>
</tbody>
</table>

Figure C.77: DM12-000 Population data
Figure C.78: DM12-000(000) Example genotype maps at 1k time-steps
Analysis of populations by genotype

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Time (steps)</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>AABB</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>AABb</td>
<td>1000</td>
<td>99.9%</td>
</tr>
<tr>
<td>AAbb</td>
<td>2000</td>
<td>99.8%</td>
</tr>
<tr>
<td>AAaB</td>
<td>3000</td>
<td>99.7%</td>
</tr>
<tr>
<td>AAaBB</td>
<td>4000</td>
<td>99.6%</td>
</tr>
<tr>
<td>AAAAB</td>
<td>5000</td>
<td>99.5%</td>
</tr>
<tr>
<td>AAAAAB</td>
<td>6000</td>
<td>99.4%</td>
</tr>
<tr>
<td>AAAAABB</td>
<td>7000</td>
<td>99.3%</td>
</tr>
<tr>
<td>AAAAABBB</td>
<td>8000</td>
<td>99.2%</td>
</tr>
<tr>
<td>AAAAABBBB</td>
<td>9000</td>
<td>99.1%</td>
</tr>
<tr>
<td>AAAAABBBB</td>
<td>10000</td>
<td>99.0%</td>
</tr>
</tbody>
</table>

Average for all runs of percentage of maximum population for each genotype

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Time (steps)</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>AABB</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>AABb</td>
<td>1000</td>
<td>99.9%</td>
</tr>
<tr>
<td>AAbb</td>
<td>2000</td>
<td>99.8%</td>
</tr>
<tr>
<td>AAaB</td>
<td>3000</td>
<td>99.7%</td>
</tr>
<tr>
<td>AAaBB</td>
<td>4000</td>
<td>99.6%</td>
</tr>
<tr>
<td>AAAAB</td>
<td>5000</td>
<td>99.5%</td>
</tr>
<tr>
<td>AAAABB</td>
<td>6000</td>
<td>99.4%</td>
</tr>
<tr>
<td>AAAAAB</td>
<td>7000</td>
<td>99.3%</td>
</tr>
<tr>
<td>AAAAABB</td>
<td>8000</td>
<td>99.2%</td>
</tr>
<tr>
<td>AAAAABBB</td>
<td>9000</td>
<td>99.1%</td>
</tr>
<tr>
<td>AAAAABBBB</td>
<td>10000</td>
<td>99.0%</td>
</tr>
</tbody>
</table>

Instances of genotypes

<table>
<thead>
<tr>
<th>Time (steps)</th>
<th>Instances</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>9</td>
<td>9.1%</td>
</tr>
<tr>
<td>1000</td>
<td>9</td>
<td>9.1%</td>
</tr>
<tr>
<td>2000</td>
<td>9</td>
<td>9.1%</td>
</tr>
<tr>
<td>3000</td>
<td>9</td>
<td>9.1%</td>
</tr>
<tr>
<td>4000</td>
<td>9</td>
<td>9.1%</td>
</tr>
<tr>
<td>5000</td>
<td>9</td>
<td>9.1%</td>
</tr>
<tr>
<td>6000</td>
<td>9</td>
<td>9.1%</td>
</tr>
<tr>
<td>7000</td>
<td>9</td>
<td>9.1%</td>
</tr>
<tr>
<td>8000</td>
<td>9</td>
<td>9.1%</td>
</tr>
<tr>
<td>9000</td>
<td>9</td>
<td>9.1%</td>
</tr>
<tr>
<td>10000</td>
<td>9</td>
<td>9.1%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time (steps)</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>9.1%</td>
</tr>
<tr>
<td>1000</td>
<td>9.1%</td>
</tr>
<tr>
<td>2000</td>
<td>9.1%</td>
</tr>
<tr>
<td>3000</td>
<td>9.1%</td>
</tr>
<tr>
<td>4000</td>
<td>9.1%</td>
</tr>
<tr>
<td>5000</td>
<td>9.1%</td>
</tr>
<tr>
<td>6000</td>
<td>9.1%</td>
</tr>
<tr>
<td>7000</td>
<td>9.1%</td>
</tr>
<tr>
<td>8000</td>
<td>9.1%</td>
</tr>
<tr>
<td>9000</td>
<td>9.1%</td>
</tr>
<tr>
<td>10000</td>
<td>9.1%</td>
</tr>
</tbody>
</table>

Case: Mutation=0.10;
Figure C.80: DM12-001(000) Example genotype maps at 1k time-steps
### Analysis of populations by genotype

Source: DM12_reportID002.cav

#### Total populations of genotypes for all runs

<table>
<thead>
<tr>
<th>Allele</th>
<th>init. types</th>
<th>100</th>
<th>200</th>
<th>300</th>
<th>400</th>
<th>500</th>
<th>600</th>
<th>700</th>
<th>800</th>
<th>900</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aabb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAbb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AaBb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aabb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>abb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Average for all runs of percentage of maximum population for each genotype

<table>
<thead>
<tr>
<th>Allele</th>
<th>init. types</th>
<th>100</th>
<th>200</th>
<th>300</th>
<th>400</th>
<th>500</th>
<th>600</th>
<th>700</th>
<th>800</th>
<th>900</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aabb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAbb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AaBb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aabb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>abb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Instances of genotypes

<table>
<thead>
<tr>
<th>Allele</th>
<th>init. types</th>
<th>100</th>
<th>200</th>
<th>300</th>
<th>400</th>
<th>500</th>
<th>600</th>
<th>700</th>
<th>800</th>
<th>900</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aabb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAbb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AaBb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aabb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>abb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

Figure C.81: DM12-002 Population data

---

Case: Mutation=0.20;
Figure C.82: DM12-002(000) Example genotype maps at 1k time-steps
### Analysis of populations by genotype

**Source:** DM12_report003.csv

**Init. pop.:** 2500  
**Init. types:** 3  
**Reruns:** 20

#### Total populations of genotypes for all reruns

<table>
<thead>
<tr>
<th>Time</th>
<th>AaBB</th>
<th>aaBB</th>
<th>AaBb</th>
<th>aabb</th>
<th>AAbb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Init.</td>
<td>49000</td>
<td>47147</td>
<td>48547</td>
<td>44175</td>
<td>42974</td>
</tr>
<tr>
<td>Reruns</td>
<td>12000</td>
<td>11441</td>
<td>12200</td>
<td>11841</td>
<td>11517</td>
</tr>
<tr>
<td></td>
<td>47174</td>
<td>45766</td>
<td>44175</td>
<td>42974</td>
<td>41953</td>
</tr>
<tr>
<td></td>
<td>4656</td>
<td>4972</td>
<td>4915</td>
<td>5330</td>
<td>5692</td>
</tr>
<tr>
<td></td>
<td>4493</td>
<td>6129</td>
<td>6736</td>
<td>7391</td>
<td>7954</td>
</tr>
<tr>
<td></td>
<td>6852</td>
<td>8221</td>
<td>8527</td>
<td>9337</td>
<td>8452</td>
</tr>
</tbody>
</table>

#### Average for all reruns of percentage of maximum population for each genotype

<table>
<thead>
<tr>
<th>Time</th>
<th>aabb</th>
<th>aaBB</th>
<th>AaBb</th>
<th>AaBB</th>
<th>AAbb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Init.</td>
<td>20000</td>
<td>25000</td>
<td>25000</td>
<td>25000</td>
<td>25000</td>
</tr>
<tr>
<td>Reruns</td>
<td>10000</td>
<td>10000</td>
<td>10000</td>
<td>10000</td>
<td>10000</td>
</tr>
<tr>
<td></td>
<td>10000</td>
<td>10000</td>
<td>10000</td>
<td>10000</td>
<td>10000</td>
</tr>
<tr>
<td></td>
<td>10000</td>
<td>10000</td>
<td>10000</td>
<td>10000</td>
<td>10000</td>
</tr>
</tbody>
</table>

#### Instances of genotypes

<table>
<thead>
<tr>
<th>Time</th>
<th>Average No. of genotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Init.</td>
<td>3752</td>
</tr>
<tr>
<td>Reruns</td>
<td>40000</td>
</tr>
</tbody>
</table>

**Case:** Mutation=0.30;

---

**Figure C.83:** DM12-003 Population data

---

294
Figure C.84: DM12-003(000) Example genotype maps at 1k time-steps
Analysis of populations by genotype

Source: DM12_reportF004.csv

init. pop: 2500
init. types: 3
Repeats: 20

Total populations of genotypes for all reruns

<table>
<thead>
<tr>
<th>Alleles</th>
<th>Time</th>
<th>Init. pop.</th>
<th>50000</th>
<th>25000</th>
<th>12500</th>
<th>6250</th>
<th>3125</th>
<th>1562</th>
<th>781</th>
<th>391</th>
<th>196</th>
<th>98.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>AABB</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
</tr>
<tr>
<td>AABB</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
</tr>
<tr>
<td>AABB</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
</tr>
<tr>
<td>AABB</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
</tr>
</tbody>
</table>

Average for all reruns of percentage of maximum population for each genotype

<table>
<thead>
<tr>
<th>Alleles</th>
<th>Time</th>
<th>Init. pop.</th>
<th>50000</th>
<th>25000</th>
<th>12500</th>
<th>6250</th>
<th>3125</th>
<th>1562</th>
<th>781</th>
<th>391</th>
<th>196</th>
<th>98.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>AABB</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
</tr>
<tr>
<td>AABB</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
</tr>
<tr>
<td>AABB</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
</tr>
<tr>
<td>AABB</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
</tr>
</tbody>
</table>

Figure C.85: DM12-004 Population data
Figure C.86: DM12-004(000) Example genotype maps at 1k time-steps
## Analysis of populations by genotype

**Source:** DM12_reportF005.csv OK

**Init. pop.:** 2500

**Init. types:** 3

**Reruns:** 20

### Total populations of genotypes for all reruns

<table>
<thead>
<tr>
<th>Alleles</th>
<th>time</th>
<th>AABb</th>
<th>AaBB</th>
<th>Aabb</th>
<th>aaBb</th>
<th>aabb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1000</td>
<td>2000</td>
<td>3000</td>
<td>4000</td>
<td>5000</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>6000</td>
<td>6000</td>
<td>6000</td>
<td>6000</td>
<td>6000</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>6000</td>
<td>6000</td>
<td>6000</td>
<td>6000</td>
<td>6000</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>6000</td>
<td>6000</td>
<td>6000</td>
<td>6000</td>
<td>6000</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>6000</td>
<td>6000</td>
<td>6000</td>
<td>6000</td>
<td>6000</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>6000</td>
<td>6000</td>
<td>6000</td>
<td>6000</td>
<td>6000</td>
</tr>
</tbody>
</table>

### Average for all reruns of percentage of maximum population for each genotype

<table>
<thead>
<tr>
<th>Alleles</th>
<th>Time</th>
<th>aabb</th>
<th>aABb</th>
<th>aABB</th>
<th>Aabb</th>
<th>AABb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>98.0</td>
<td>93.0</td>
<td>97.0</td>
<td>98.0</td>
<td>94.0</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>98.0</td>
<td>93.0</td>
<td>97.0</td>
<td>98.0</td>
<td>94.0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>98.0</td>
<td>93.0</td>
<td>97.0</td>
<td>98.0</td>
<td>94.0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>98.0</td>
<td>93.0</td>
<td>97.0</td>
<td>98.0</td>
<td>94.0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>98.0</td>
<td>93.0</td>
<td>97.0</td>
<td>98.0</td>
<td>94.0</td>
</tr>
</tbody>
</table>

### Instances of genotypes

| Instances | Time | 0 | 1000 | 2000 | 3000 | 4000 | 5000 | 6000 | 7000 | 8000 | 9000 | 10000 | 11000 | 12000 | 13000 | 14000 | 15000 | 16000 | 17000 | 18000 | 19000 | 20000 |
|-----------|------|---|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| aabb      |     | 0 | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 1.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  |
| aABb      |     | 0 | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 1.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  |
| aABB      |     | 0 | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 1.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  |
| Aabb      |     | 0 | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 1.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  |
| AABb      |     | 0 | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 1.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  | 0.0  |

### Case: Mutation=0.50

Figure C.87: DM12-005 Population data
Figure C.88: DM12-005(000) Example genotype maps at 1k time-steps
C.5 Simulations of Dobzhansky-Müller with predators

This Appendix gives fuller details of the predation simulations described in Section 5.5 on page 135.

C.5.1 Predation as the $AB$ excluder (PR00)

This simulation is discussed in Section 5.5.1 on page 135.

C.5.1.1 Predation as the $AB$ excluder (PR00): input

The input for this simulation is shown in Table C.24 on page 301.

C.5.1.2 Predation as the $AB$ excluder (PR00): results

The results are shown in Figures C.89 to C.96. In each case the first figure shows the percentage of the population with each genotype and the second shows the development of genotype groups as a map of the lattice.
//Sympatria input file
Computer: Toshiba Satellite T110-107 Laptop;

//TITLES
Purpose: From Aabb and aaBb with no extra mutation AB predated with and without AM;

//LATTICE
Rows: 50;
Cols: 50;

//INITIAL SETUP Aabb and aaBb
InitialSet: LeftRight FFFFFFFFFFFEFFFFFFFFFF
            FFF7FFFFFFFFFFFFF7FFFFFFFFFF;

//BASE DATA
MateRange: 1;
SeedRange: 1;
MatureAge: 1;
Seeds: 1;
Recomb: 0.00; //NONE
CrossOvers: 0.00; //NONE
Mutation: 0.00; //NONE

//PHENOTYPE FATALITY (a fatal combination of mutations)
//NONE

//ENVIRONMENTAL DEATH
EnvDeathRate: 0.01; //default 0.01

//PREDATORS
PredationRate: 0.50, 1.00
PredAttractLoci: 8x000000000080000000; //loci tested
PredAttractGenes: 8x000000000000000000; //attractive gene pattern
PredDeterLoci: 8x000000000000000000; //df ZERO
PredDeterGenes: 8x000000000000000000; //df ZERO

//PHENOTYPE COMPATIBILITY
//NONE

//ASSORTATIVE MATING
AssortativeLoci: 8x000000000000000000; //OFF and ON

//CYCLES
Generations: 100000;
Stops: 20;
ReRuns: 28;
RandomSeed: 999;

//rptF: Nr of Family types v Time
ReportFamilies: YES;
ReportLoci: 0xFFFFFFFFFFFFFFFF;

//Colour mapping
GeneImages: YES;
CaptureReruns: 1;
RedMask: 0x000000000000000000;
GreenMask: 0x000000000000000000;
BlueMask: 0x000000000000000000;
HapCombination: SPLIT;

Table C.24: PR00 input data
### Analysis of populations by genotype

Source: PR00_reportF000.csv

<table>
<thead>
<tr>
<th>Time step</th>
<th>Initiator population</th>
<th>Initiator types</th>
<th>Reruns</th>
<th>Total populations of genotypes for all reruns</th>
<th>Average No of genotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>10000</td>
<td>2500</td>
<td>2</td>
<td>20</td>
<td>2500</td>
<td>2500</td>
</tr>
<tr>
<td>50000</td>
<td>101</td>
<td>106</td>
<td>58</td>
<td>1001</td>
<td>101.00</td>
</tr>
<tr>
<td>90000</td>
<td>100</td>
<td>106</td>
<td>58</td>
<td>1000</td>
<td>100.00</td>
</tr>
</tbody>
</table>

#### Instances of genotypes

<table>
<thead>
<tr>
<th>Instance</th>
<th>Average No of genotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2.50</td>
</tr>
</tbody>
</table>

#### Figure C.89: PR00-000 Population data
Figure C.90: PR00-000(000) Example genotype maps at 5k time-steps
Analysis of populations by genotype

Init. pop: 2500
Init. types: 2
Reruns: 20

Total populations of genotypes for all reruns

<table>
<thead>
<tr>
<th>Time</th>
<th>Alleles</th>
<th>AABB</th>
<th>aaBB</th>
<th>AABb</th>
<th>AaBB</th>
<th>AaBb</th>
<th>aaBb</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>11013</td>
<td>6157</td>
<td>8393</td>
<td>1963</td>
<td>1154</td>
<td>681</td>
<td>444</td>
</tr>
<tr>
<td>2500</td>
<td>2982</td>
<td>1206</td>
<td>8594</td>
<td>4805</td>
<td>250</td>
<td>108</td>
<td>55</td>
</tr>
<tr>
<td>2000</td>
<td>2247</td>
<td>2200</td>
<td>2247</td>
<td>2000</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2500</td>
<td>2284</td>
<td>1223</td>
<td>2284</td>
<td>2284</td>
<td>2284</td>
<td>2284</td>
<td>2284</td>
</tr>
<tr>
<td>2000</td>
<td>2284</td>
<td>2284</td>
<td>2284</td>
<td>2284</td>
<td>2284</td>
<td>2284</td>
<td>2284</td>
</tr>
</tbody>
</table>

Average for all reruns of percentage of maximum population for each genotype

AABB: 0.0% 22.0% 12.3% 7.3% 3.0% 2.3% 1.4% 0.9% 1.0% 1.1% 0.7% 0.9% 0.3% 0.3% 0.1% 0.1% 0.0% 0.0% 0.0% 0.0%
aaBB: 50.0% 3.4% 2.5% 1.3% 0.2% 0.2% 0.2% 0.2% 0.1% 0.1% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0%
aABb: 0.0% 31.6% 31.6% 41.5% 42.5% 41.5% 41.5% 41.5% 41.5% 41.5% 41.5% 41.5% 41.5% 41.5% 41.5% 41.5% 41.5% 41.5% 41.5% 41.5%
AaBB: 50.0% 5.7% 3.5% 1.8% 1.0% 0.5% 0.3% 0.2% 0.1% 0.1% 0.1% 0.1% 0.1% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0%
aABB: 0.0% 34.2% 41.5% 45.9% 48.9% 51.8% 52.1% 52.2% 52.1% 52.2% 52.1% 52.2% 52.2% 52.1% 52.2% 52.1% 52.2% 52.1% 52.2% 52.1%
AaBb: 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0%
aaBb: 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0%
aABB: 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0%
AAbb: 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0% 0.0%
AABb: 100.0% 99.1% 99.0% 99.0% 99.0% 99.0% 99.0% 99.0% 99.0% 99.0% 99.0% 99.0% 99.0% 99.0% 99.0% 99.0% 99.0% 99.0% 99.0% 99.0%

Instances of genotypes

<table>
<thead>
<tr>
<th>Time</th>
<th>Instances</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>300</td>
<td>3.00</td>
</tr>
<tr>
<td>2500</td>
<td>2500</td>
<td>2.00</td>
</tr>
<tr>
<td>2500</td>
<td>2500</td>
<td>2.00</td>
</tr>
<tr>
<td>2500</td>
<td>2500</td>
<td>2.00</td>
</tr>
<tr>
<td>2500</td>
<td>2500</td>
<td>2.00</td>
</tr>
</tbody>
</table>

Case: PredationRate=0.50; AssortativeIndex=4;

Figure C.91: PR00-001 Population data
Figure C.92: PR00-001(000) Example genotype maps at 5k time-steps
Analysis of populations by genotype

Source: PR00_reportF002.csv OK

Init. pop. 2500
Init. types 2
Reruns 20

Total populations of genotypes for all reruns

<table>
<thead>
<tr>
<th>Alleles</th>
<th>AABB</th>
<th>AAbb</th>
<th>Aabb</th>
<th>Total populations of genotypes for all reruns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reruns</td>
<td>Init. types</td>
<td>Average</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>5000</td>
<td>10000</td>
<td>15000</td>
<td>20000</td>
</tr>
<tr>
<td>Time</td>
<td>0</td>
<td>5000</td>
<td>10000</td>
<td>15000</td>
</tr>
<tr>
<td>AABB</td>
<td>0</td>
<td>19407</td>
<td>24172</td>
<td>27172</td>
</tr>
<tr>
<td>AAbb</td>
<td>25000</td>
<td>8519</td>
<td>56918</td>
<td>52761</td>
</tr>
<tr>
<td>AaBb</td>
<td>0</td>
<td>6992</td>
<td>56681</td>
<td>52777</td>
</tr>
</tbody>
</table>

Average for all reruns of percentage of maximum population for each genotype

| Time    | 0    | 5000  | 10000 | 15000 | 20000 | 25000 | 30000 | 35000 | 40000 | 45000 | 50000 | 55000 | 60000 | 65000 | 70000 | 75000 | 80000 | 85000 | 90000 | 95000 | 100000 |
| AABB    | 0.0% | 38.8% | 48.4% | 54.2% | 58.7% | 64.6% | 66.8% | 70.4% | 75.8% | 78.6% | 78.9% | 79.6% | 79.6% | 81.6% | 81.7% | 81.9% | 82.9% | 83.2% | 82.9% | 83.3% | 84.4% |
| AAbb    | 50.0% | 17.2% | 21.5% | 13.2% | 12.4% | 11.5% | 10.3% | 7.8% | 7.5% | 7.1% | 7.3% | 7.0% | 6.9% | 6.7% | 6.1% | 6.1% | 6.5% | 6.8% | 6.3% | 6.2% | 5.7% |
| AaBb    | 0.0% | 13.4% | 11.4% | 10.7% | 9.7% | 8.3% | 6.1% | 5.0% | 4.0% | 3.4% | 3.3% | 3.9% | 3.7% | 3.7% | 3.4% | 3.6% | 3.6% | 3.4% | 3.4% | 3.6% | 3.8% |

Instances of genotypes

| Time    | 0    | 5000  | 10000 | 15000 | 20000 | 25000 | 30000 | 35000 | 40000 | 45000 | 50000 | 55000 | 60000 | 65000 | 70000 | 75000 | 80000 | 85000 | 90000 | 95000 | 100000 |
| Instances | 2 | 104 | 110 | 100 | 98 | 95 | 95 | 81 | 78 | 75 | 75 | 68 | 62 | 60 | 57 | 54 | 54 | 49 | 46 | 46 | 46 |

Average No of genotypes

Case: PredationRate=1.00; AssortativeIndex=0;

Figure C.93: PR00-002 Population data
Figure C.94: PR00-002(000) Example genotype maps at 5k time-steps
### Analysis of populations by genotype

**Source:** PR00_reportF003.csv

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Init. pop.</th>
<th>Init. types</th>
<th>Reruns</th>
</tr>
</thead>
<tbody>
<tr>
<td>AABB</td>
<td>0.00</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AABb</td>
<td>0.00</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AABB</td>
<td>0.00</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Aabb</td>
<td>0.00</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>aaBb</td>
<td>0.00</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>aabb</td>
<td>0.00</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

#### Total populations of genotypes for all reruns

<table>
<thead>
<tr>
<th>Case</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100.0%</td>
</tr>
<tr>
<td>2</td>
<td>50.0%</td>
</tr>
<tr>
<td>3</td>
<td>50000</td>
</tr>
</tbody>
</table>

#### Average for all reruns of percentage of maximum population for each genotype

<table>
<thead>
<tr>
<th>Case</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100.0%</td>
</tr>
<tr>
<td>2</td>
<td>50.0%</td>
</tr>
<tr>
<td>3</td>
<td>50000</td>
</tr>
</tbody>
</table>

#### Instances of genotypes

<table>
<thead>
<tr>
<th>Case</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100.0%</td>
</tr>
<tr>
<td>2</td>
<td>50.0%</td>
</tr>
<tr>
<td>3</td>
<td>50000</td>
</tr>
</tbody>
</table>

---

**Case:** PredationRate=1.00; AssortativeIndex=4;

---

**Figure C.95:** PR00-003 Population data
Figure C.96: PR00-003(000) Example genotype maps at 5k time-steps
C.5.2 Predation to deplete the \textit{aabb} forms (PR01)

This simulation is discussed in Section 5.5.2 on page 137.

C.5.2.1 Predation to deplete the \textit{aabb} forms (PR01): input

The input for this simulation is shown in Table C.25 on page 310.

```
//Sympatria input file
Computer: Toshiba Satellite T110-107 Laptop;
//TITLES
Purpose: aabb with 0.1 mutation aabb predation AB exclusion but no crossover;
//LATTICE
Rows: 50;
Cols: 50;
//INITIAL SETUP aabb with Aabb and aaBb in 2x2 corners
InitialSet: ABcorners 0x0000000000000001 0x0000000000000002 2;
//BASE DATA
MateRange: 8;
SeedRange: 1;
MatureAge: 1;
Seeds: 1;
Recomb: 0.00;
CrossOvers: 0.00;
Mutation: 0.00;
//PHENOTYPE FATALITY (a fatal combination of mutations)
PhenotypeFatal: 0x0000000000000003;
//PHENOTYPE COMPATIBILITY
PhenotypeLoci: 0x0000000000000003;
PhenoTolerance: 1;
//ASSORTATIVE MATING
AssortativeLoci: 0x0000000000000003;
AssortativeIndex: 8;
//PREDATION
PredationRate: 0.00, 0.05, 0.10;
PredAttractLoci: 0x0000000000000000; //loci of interest
PredAttractGenes: 0x0000000000000000; //aabb genes
//ENVIRONMENTAL DEATH
EnvDeathRate: 0.01; //default 0.01
//CYCLES
Generations: 500;
Stops: 25;
ReRuns: 20;
RandomSeed: 999;
//rptF: Nr of Family types v Time
ReportFamilies: YES;
ReportLoci: 0xFFFFFFFFFFFFFFFF;
//Colour mapping
GeneImages: Yes;
CaptureReruns: 2;
RedMask: 0x0000000000000001;
GreenMask: 0x0000000000000002;
BlueMask: 0x0000000000000004;
HapCombination: SPLIT;
//Movie maker
MakeMovie: YES;
ImageDwell: 0.20; //seconds per frame
KeepImages: Yes;
```

Table C.25: PR01 input data

The predation death rate ($\chi P$) is progressively increase through 0%, 5% to 10%.
C.5.2.2 Predation to deplete the \textit{aabb} forms (PR01): results

The genotype maps in Figures C.97 to C.99 show how the spread of the \textit{Aabb} and \textit{aaBb} genotypes improves as the predation rate for the \textit{aabb} type is increased.
Figure C.97: PR01-000(000) Early genotype maps for $D_P = 0\%$
Figure C.98: PR01-001(000) Early genotype maps for $D_P = 5\%$
Figure C.99: PR01-002(000) Early genotype maps for $D_P = 10\%$
C.5.3 Varying the general death and predation rates (PR18-21)

These simulations are discussed in Section 5.5.3 on page 141.

C.5.3.1 Varying the general death and predation rates (PR18-21): input

The input file for these simulations in Tables C.26 to C.29.

//Sympatria input file
Computer: Toshiba Satellite T110-107 Laptop;
//TITLES
Purpose: aabb predation De 0.5 and AB exclusion;
//LATTICE
Rows: 50;
Cols: 50;
//INITIAL SETUP aabb with Aabb and aaBb in corners
InitialSet: ABcorners 0x0000000000000001 0x0000000000000002 2;
//BASE DATA
MateRange: 8;
SeedRange: 1;
MatureAge: 1;
Seeds: 1;
Recomb: 0.00;
Mutation: 0.00;
//ENVIRONMENT
EnvDeathRate: 0.005;
//PHENOTYPE FATALITY (a fatal combination of mutations)
PhenotypeFatal: 0x0000000000000003;
//PHENOTYPE COMPATIBILITY
PhenotypeLoci: 0x0000000000000003;
PhenoTolerance: 1;
//ASSORTATIVE MATING
AssortativeLoci: 0x0000000000000003;
AssortativeIndex: 8;
//PREDATION
PredationRate: 0.000, 0.001, 0.002, 0.003, 0.004, 0.005, 0.006, 0.007, 0.008, 0.009, 0.010;
PredAttractLoci: 0x0000000000000003; //loci of interest
PredAttractGenes: 0x0000000000000000; //aabb genes
//CYCLES
Generations: 2000;
Stops: 20;
ReRuns: 50;
RandomSeed: 999;
//rptF: Nr of Family types v Time
ReportFamilies: YES;
ReportLoci: 0xFFFFFFFFFFFFFFFF;
//Colour mapping
GeneImages: Yes;
CaptureReruns: 2;
RedMask: 0x0000000000000001;
GreenMask: 0x0000000000000002;
BlueMask: 0x0000000000000004;
HapCombination: SPLIT;

Table C.26: PR18 input data
//Sympatria input file
Computer: Toshiba Satellite T110-107 Laptop;
//TITLES
Purpose: aabb predation De 4% to 40% and AB exclusion;
//LATTICE
Rows: 50;
Cols: 50;
//INITIAL SETUP aabb with Aabb and aaBb in corners
InitialSet: ABcorners 0x0000000000000001 0x0000000000000002 2;
//BASE DATA
MateRange: 8;
SeedRange: 1;
MatureAge: 1;
Seeds: 1;
Recomb: 0.00;
CrossOvers: 0.00;
Mutation: 0.00;
//ENVIRONMENT
EnvDeathRate: 0.04, 0.10, 0.20, 0.40;
//PHENOTYPE FATALITY (a fatal combination of mutations)
PhenotypeFatal: 0x0000000000000003;
//PHENOTYPE COMPATIBILITY
PhenotypeLoci: 0x0000000000000003;
PhenoTolerance: 1;
//ASSORTATIVE MATING
AssortativeLoci: 0x0000000000000003;
AssortativeIndex: 8;
//PREDATION
PredationRate: 0.000, 0.001, 0.002, 0.003, 0.004, 0.005, 0.006, 0.007, 0.008, 0.009, 0.010;
PredAttractLoci: 0x0000000000000003; //loci of interest
PredAttractGenes: 0x0000000000000000; //aabb genes
//CYCLES
Generations: 2000;
Stops: 20;
ReRuns: 50;
RandomSeed: 999;
//rptF: Nr of Family types v Time
ReportFamilies: YES;
ReportLoci: 0xFFFFFFFFFFFFFFFF;
//Colour mapping
GeneImages: Yes;
CaptureReruns: 2;
RedMask: 0x0000000000000001;
GreenMask: 0x0000000000000002;
BlueMask: 0x0000000000000004;
HapCombination: SPLIT;

Table C.27: PR19 input data
//Sympatria input file
Computer: Toshiba Satellite T110-107 Laptop;

//TITLES
Purpose: aabb predation D0 0% and 2% with AB exclusion;

//LATTICE
Rows: 50;
Cols: 50;

//INITIAL SETUP aabb with Aabb and aaBb in corners
InitialSet: ABcorners 0x0000000000000001 0x0000000000000002 2;

//BASE DATA
MateRange: 8;
SeedRange: 1;
MatureAge: 1;
Seeds: 1;
Recomb: 0.00;
CrossOvers: 0.00;
Mutation: 0.00;

//ENVIRONMENT
EnvDeathRate: 0.0, 0.02;

//PHENOTYPE FATALITY (a fatal combination of mutations)
PhenotypeFatal: 0x0000000000000003;

//PHENOTYPE COMPATIBILITY
PhenotypeLoci: 0x0000000000000003;
PhenoTolerance: 1;

//ASSORTATIVE MATING
AssortativeLoci: 0x0000000000000003;
AssortativeIndex: 8;

//PREDATION
PredationRate: 0.000, 0.001, 0.002, 0.003, 0.004, 0.005, 0.006, 0.007, 0.008, 0.009, 0.010;
PredAttractLoci: 0x0000000000000003; //loci of interest
PredAttractGenes: 0x0000000000000000; //aabb genes

//CYCLES
Generations: 2000;
Stops: 20;
ReRuns: 50;
RandomSeed: 999;

//rptF: Nr of Family types v Time
ReportFamilies: YES;
ReportLoci: 0xFFFFFFFFFFFFFFFF;

//Colour mapping
GeneImages: Yes;
CaptureRuns: 2;
RedMask: 0x0000000000000001;
GreenMask: 0x0000000000000002;
BlueMask: 0x0000000000000004;
HapCombination: SPLIT;

Table C.28: PR20 input data
Table C.29: PR21 input data
C.5.3.2 Varying the general death and predation rates (PR18-21): results

The results described here were taken in part from runs PR18, 19, 20 and 21 although parts of some of these runs duplicated each other. The results for each value of \( D_0 \) were taken from the following runs:

PR18 \( D_0 = 0.5\% \)
PR19 \( D_0 = 4\%, 10\%, 20\%, \) and \( 40\% \)
PR20 \( D_0 = 0\% \) and \( 2\% \)
PR21 \( D_0 = 1\% \) and \( 3\% \)

Because of the large volumes of output data (99 cases each run 50 times) the populations means were collated using a small C++ routine and then plotted using a spreadsheet.

<table>
<thead>
<tr>
<th>( D_0 )</th>
<th>0.0%</th>
<th>0.1%</th>
<th>0.2%</th>
<th>0.3%</th>
<th>0.4%</th>
<th>0.5%</th>
<th>0.6%</th>
<th>0.7%</th>
<th>0.8%</th>
<th>0.9%</th>
<th>1.0%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0%</td>
<td>0</td>
<td>2.3</td>
<td>5.6</td>
<td>13.4</td>
<td>26.3</td>
<td>40.6</td>
<td>57.3</td>
<td>89.1</td>
<td>118.7</td>
<td>138.2</td>
<td>204.1</td>
</tr>
<tr>
<td>0.5%</td>
<td>1.0</td>
<td>7.4</td>
<td>15.0</td>
<td>22.2</td>
<td>43.9</td>
<td>71.6</td>
<td>98.4</td>
<td>149.3</td>
<td>200.8</td>
<td>250.0</td>
<td>317.1</td>
</tr>
<tr>
<td>1.0%</td>
<td>1.5</td>
<td>2.8</td>
<td>7.9</td>
<td>15.7</td>
<td>34.4</td>
<td>67.9</td>
<td>93.0</td>
<td>150.7</td>
<td>226.3</td>
<td>382.4</td>
<td></td>
</tr>
<tr>
<td>2.0%</td>
<td>0.9</td>
<td>3.7</td>
<td>4.0</td>
<td>17.2</td>
<td>16.4</td>
<td>59.0</td>
<td>86.4</td>
<td>111.2</td>
<td>132.8</td>
<td>224.2</td>
<td>241.4</td>
</tr>
<tr>
<td>3.0%</td>
<td>1.2</td>
<td>0.0</td>
<td>1.7</td>
<td>13.6</td>
<td>11.7</td>
<td>32.8</td>
<td>52.6</td>
<td>46.2</td>
<td>54.1</td>
<td>140.5</td>
<td>176.7</td>
</tr>
<tr>
<td>4.0%</td>
<td>0.0</td>
<td>0.9</td>
<td>0.9</td>
<td>15.3</td>
<td>19.3</td>
<td>37.7</td>
<td>26.1</td>
<td>68.7</td>
<td>97.6</td>
<td>100.9</td>
<td>111.8</td>
</tr>
<tr>
<td>10.0%</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>2.8</td>
<td>0.0</td>
<td>22.3</td>
<td>12.4</td>
<td>48.1</td>
<td>83.6</td>
<td>48.5</td>
<td>36.5</td>
</tr>
<tr>
<td>20.0%</td>
<td>4.3</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>14.1</td>
<td>16.3</td>
<td>0.0</td>
<td>71.1</td>
<td>23.9</td>
<td>110.5</td>
</tr>
<tr>
<td>40.0%</td>
<td>2.8</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>32.5</td>
<td>56.4</td>
<td>0.0</td>
<td>33.0</td>
<td>33.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Table C.30: Mean (AAbb & aaBB) population at \( t=2000 \) (PR18-21)
Figure C.100: Mean (AAbb & aaBB) population at $t=2000$ v $D_0$ (PR18-21)


