Reproductive isolation associated with the copper tolerance locus in *Mimulus guttatus*

Submitted by Deborah L. Lloyd to the University of Exeter
as a thesis for the degree of
Doctor of Philosophy in Biological Sciences
in October 2010

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: ……..

Acknowledgements

I would like to offer my gratitude to my supervisor, Prof. Mark Macnair, for his encouragement, support and advice during this project, and to my collaborators, and in particular Kevin Wright, in the John Willis Lab at Duke University, North Carolina, for their contribution and support and for welcoming me to their laboratory to undertake some aspects of this project. I would also like to thank Profs John Bryant, Murray Grant and Nick Smirnoff for their helpful feedback on chapters of this thesis.

I also extend my thanks to Jim Chidlow for helping me so much in looking after my *Mimulus* plants, for his advice on all aspects of growing and caring for the plants and for his friendly companionship during the long hours I spent in the greenhouse.

I am hugely grateful for the invaluable encouragement and friendship I have received from my colleagues in the Fungal Group at Exeter University. I would particularly like to thank Dr Steve Bates for his understanding and support as I balanced my work for him with my PhD.

Most of all, I especially wish to thank my friends and family and especially my mother for her endless support, encouragement and understanding throughout the duration of my PhD, although thanks does not seem enough for one who has suffered the highs and lows as much as I. My heartfelt thanks also to my sister and my friends who have helped me through the hard times and forgiven me for not seeing them more often due to work. Without you all this thesis would never have been.

Abstract

The evolution of reproductive isolating barriers that prevent gene flow between species is essential to the process of speciation. One such barrier is intrinsic postzygotic isolation, which proceeds as hybrid sterility or inviability, and is commonly attributed to Dobzhansky-Muller genic incompatibilities. Here, deleterious interlocus interactions occur between incompatible alleles of complementary genes when brought together in the genome of a hybrid. Although these hybrid incompatibilities are widespread, having been identified in mammals, fish, plants and fungi, still relatively little is known about the nature of the genes involved.

In the model plant species Mimulus, a Dobzhansky-Muller incompatibility exists between two populations of the yellow monkey flower, *Mimulus guttatus*, in which the interaction between a single gene from a copper tolerant population, Copperopolis, and a small number of polymorphic genes from a second non-tolerant population, Cerig-y-drudion, results in hybrid necrosis in the F_1. Hybrid necrosis, a form of hybrid inviability with phenotypic characteristics strongly similar to those of plants responding to pathogen attack, is a common barrier preventing hybridization in plants. As well as being of interest in terms of evolution, hybrid necrosis has practical implications in plant breeding as it prevents the combining of desirable traits from related species in commercial cultivars.

In the cross between Copperopolis and Cerig-y-drudion, copper tolerance, conferred by a single major gene, and hybrid necrosis are tightly linked but the independent or synonymous nature of the gene(s) in the Copperopolis population that contribute to these two characteristics is unknown.

A key aim of this thesis was to establish the nature of the single gene in Copperopolis that contributes to hybrid necrosis with regards to its linkage to copper tolerance. The gene for hybrid necrosis was found to be tightly linked to, but discrete from, the gene controlling copper tolerance. Three candidate genes for this hybrid necrosis locus were indentified: a Jumonji-domain containing protein with probable function as a methyltransferase, a glycosyltransferase and a possible phosphatase. Interestingly, the latter two have potential functional roles in the plant immune system.

The second key aim of this thesis was to perform the first investigation into the small number of genes in the Cerig-y-drudion population that contribute to the crossing barrier. Two QTLs for hybrid necrosis were identified. One QTL on Chromosome 9 is responsible for around 20% of the hybrid necrosis whilst the second QTL on Chromosome 12 acts as an enhancer of the first QTL causing an additional 10% of necrosis. Interestingly both these QTLs contain R genes, further implicating the possible involvement of the plant immune system in this crossing barrier.

Table of Contents

	Page
Title Page	**1**
Acknowledgements	**2**
Abstract	**3**
Table of Contents	**4-10**
List of Tables and Figures	**11-17**
Author's Declaration	**18-19**
List of abbreviations	**20-22**
Chapter 1 - General Introduction	**23-64**
1.1 The BSC and other species concepts	23
1.1.1 Biogeography and Speciation	26
1.2 Reproductive Isolation	28
1.3 The Dobzhansky-Muller Model of Genic Incompatibilities	32
1.4 Forces driving the evolution of hybrid incompatibilities: Selection, drift and genomic conflict	36
1.4.1 Genetic drift	36
1.4.2 Natural Selection	36
1.4.3 Genomic Conflict	38
1.5 Examples of hybrid incompatibilities and DM genes	38
1.5.1 Hybrid lethality genes in *Drosophila*	39
1.5.2 Hybrid lethality genes in Plants	48
1.5.2.1 *Mimulus* -a model system for evolutionary studies	49
1.5.2.2 Copper tolerance and hybrid lethality in *M. guttatus*	50
1.5.2.3 Hybrid necrosis in other plant systems	54

1.5.2.4 Hybrid necrosis and the plant immune system 57

1.5.3 Hybrid lethality genes in Fish 63

1.6 Summary and Aims 63

Chapter 2 - Methods and Materials 65-84

2.1 The study species: *Mimulus guttatus* Fischer ex DC 65

2.2 Description & provenance of plants used in this research 65

2.2.1 Copperopolis 65

2.2.2 Cerig-y-Drudion 67

2.2.3 Plants ‘A’ and ‘D’ 70

2.3 Plant growth and care 70

2.3.1 Raising plants from seed 70

2.3.2 Raising plants from cuttings 70

2.3.3 Soil 71

2.3.4 Lighting/blackout 71

2.3.5 Temperature 72

2.3.6 Watering and feeding 72

2.3.7 Seed storage 73

2.3.8 Pest management 73

2.4 Glasshouse-based Experimental Techniques 73

2.4.1 Phenotyping 73

2.4.1.1 Phenotyping for Copper Tolerance 73

2.4.1.2 Phenotyping for Hybrid Necrosis 74

2.4.2 Crosses/pollination 75

2.4.3 Tissue Sampling 76

2.5 I.T. Resources 76

2.5.1 Molecular IT Tools 77

2.5.1.1 Markers 77

2.5.1.2 Whole-gene sequence amplification intended for sequencing 77

2.5.2 Bioinformatics IT Tools 77

2.6 Molecular techniques 78

2.6.1 DNA Extraction 78

2.6.1.1 The CTAB method 78

2.6.1.2 The QIAGEN DNeasy Kit 80

2.6.2 Primers 80

2.6.2.1 Primer Design 80

2.6.2.2 Primer Preparation 80

2.6.3 PCR 81

2.6.3.1 Preparing the Master Mix 81

2.6.3.2 PCR Programme 82

2.6.4 Gel Electrophoresis 82

2.6.5 Genotyping 83

2.7 Statistical Analysis 84

Chapter 3 - Characterising the phenotype of hybrid necrosis in crosses between the Cerig and Copperopolis populations of *Mimulus guttatus* **85-101**

3.1 Introduction 85

3.2 Methods 89

3.2.1 Phenotyping for hybrid necrosis 89

3.2.2 Establishing polymorphism for hybrid necrosis genes in the Cerig population 89

3.2.3 Phenotyping for Copper tolerance 90

3.2.4 Testing for Temperature/Environmental effects on the degree of hybrid necrosis 90

3.3 Results 90

3.3.1 The hybrid necrosis phenotype 90

3.3.2 Polymorphism for hybrid necrosis genes in the Cerig population 92

3.3.2.1 Phenotypic variation in 14 Cerig individuals 92

3.3.2.2 Testing for homogeneity between reciprocal crosses 95

3.3.2.3 Testing for homogeneity amongst crosses within each group 96

3.3.2.4 Differential necrosis depending on the Cerig parent 97

3.3.3 Hybrid Necrosis and Temperature 98

3.5 Discussion 99

3.6 Conclusion 101

Chapter 4 - Identifying the genomic region containing the hybrid necrosis (*NEC*) locus in Copperopolis **102-126**

4.1 Introduction 102

4.2 Methods and Materials 104

4.2.1 Identifying Recombinants 104

4.2.1.1 Phenotyping for Copper tolerance 104

4.2.1.2 Genotyping 104

4.2.2 Selection of Cerig individuals to use in testcrosses with Recombinants 106

4.2.3 Crosses between Cerigs and recombinants 106

4.2.4 Phenotyping for hybrid necrosis 107

4.2.5 Calculating Recombination Frequencies 107

4.3 Results 107

4.3.1 Testing the linkage between the *T* and *NEC* loci 107

4.3.1.1 Testing for homogeneity amongst repeat crosses 110

4.3.1.2 Testing for a parent-of-origin effect 111

4.3.1.3 Testing the effect of a TR or NTR parent on the rate of hybrid necrosis 111

4.3.1.4 Identifying recombinants between the *T* and *NEC* loci 113

4.3.2 Identifying T locus markers and the recombination window containing NEC 117

4.4 Discussion 123

4.5 Conclusion 126

Chapter 5 - Bioinformatics and fine-mapping of the recombination interval containing the hybrid necrosis (*NEC*) locus in Copperopolis **127-151**

5.1 Introduction 127

5.2 Methods 130

5.2.1 Fine-mapping of the recombination window containing *NEC* 130

5.2.1.1 Provenance of plants 131

5.2.1.2 Primer design for new markers 131

5.2.1.3 Identification of informative markers 133

5.2.2 Bioinformatics 137

5.3 Results 138

5.3.1 Fine mapping 138

5.3.2 Bioinformatics 141

5.4 Discussion 144

5.5 Conclusion 151

Chapter 6 - QTL mapping of the genes in the Cerig population that interact with the *NEC* gene in Copperopolis to cause hybrid necrosis **152-189**

6.1 Introduction 152

6.2 Methods and Materials 156

6.2.1 QTL analysis and Bulked Segregant Analysis 156

6.2.2 Generating a segregating F_2 population for BSA 157

6.2.3 Phenotyping the segregating F_2 population for hybrid necrosis 157

6.2.4 Screening for markers 159

6.2.5 Data Analysis Techniques 160

6.2.5.1 Fisher's exact test in BSA 160

6.2.5.2 Sequential Bonferroni correction 160

6.2.5.3 Arcsine transformation 160

6.2.5.4 One way Analysis of Variance (ANOVA) 162

6.3 Results 162

6.3.1 Testing replicate F_2 x Cop crosses for homogeneity 162

6.3.2 Hybrid necrosis in the segregating F_2 population 164

6.3.3 Screening for polymorphic markers in the parent plants 165

6.3.4 Bulked Segregant Analysis 165

6.3.5 Marker and QTL analysis 169

6.4 Discussion 182

6.5 Conclusion 188

Chapter 7 - General Discussion **190-198**

7.1 The single hybrid necrosis locus in Copperopolis 190

7.2 Identifying hybrid necrosis QTLs in Cerig 193

7.3 Future Work 197

7.4 Final Conclusion 198

Appendices **199-209**

Appendix 4.1 Raw data for hybrid necrosis from crosses between Cerig 10 or Cerig 35 and the tolerant (TR) and non-tolerant (NTR) recombinants 199-201

Appendix 6.1 Chi-squared values between replicates in the F_2 x Cop crosses 202-205

Appendix 6.2 Single Factor/ One Way ANOVA results for each of the three markers in Cerig that co-segregate with the hybrid necrosis phenotype 206-207

Appendix 7.1 Maps of the Mimulus genome showing the size (cM) of the fourteen chromosomes and the locations of the MgSTS markers screen 208-209

References **210-228**

List of Tables and Figures

Tables

Chapter 1 **Page**

Table 1.1 A summary of the main Species Concepts 24-25

Table 1.2 Definitions for the three modes of speciation based on biogeography 27

Table 1.3 Reproductive Isolating Barriers 29-30

Table 1.4 Examples of genes identified as contributing to Dobzhansky-Muller incompatibility interactions leading to hybrid dysfunctions 40-45

Chapter 2

Table 2.1 Specifications of plant feed 72

Chapter 3

Table 3.1 Copper tolerance and hybrid necrosis phenotypes in offspring from crosses between different Cerig parents (non-tolerant) and plant D (tolerant) 93

Table 3.2 Raw data on the degree of hybrid necrosis from crosses between Cerig and Cop plants. 94

Table 3.3 Chi-squared (χ^2) values for the difference in the degree of hybrid necrosis in reciprocal crosses between Cerigs and Cop. 96

Table 3.4 The degree of hybrid necrosis in crosses between Cop and either Cer10, Cer33, Cer34 or Cer35 96

Table 3.5 Chi-squared (χ^2) values for testing for homogeneity amongst crosses within each Cerig x Cop group. 97

Table 3.6 Chi-squared (χ^2) values for testing for homogeneity in the degree of hybrid necrosis amongst crosses with data from each Cerig being omitted sequentially. 98

Chapter 4

Table 4.1 A summary of the results of the marker screening in plants A and D and in D x A F_1 progeny as conducted by Kevin Wright at Duke University to identify markers linked to copper tolerance. 105

Table 4.2 The number of replicates scored for hybrid necrosis for each recombinant (between marker e242 and the T locus) parent plant in crosses with Cer10 and the range in the degree of hybrid necrosis within these replicates. 108

Table 4.3 Chi-squared (χ^2) values for comparing the degree of hybrid necrosis in replicate crosses for each tolerant recombinant. 110

Table 4.4 Chi-squared (χ^2) values for comparing the hybrid necrosis data from replicate crosses with each non-tolerant recombinant 110

Table 4.5 Chi-squared (χ^2) values for comparing the mean hybrid necrosis in reciprocal crosses between Cer10 and grouped or individual TRs and NTRs. 112

Table 4.6 Pooled data for hybrid necrosis in TR and NTR crosses with Cer10 113

Table 4.7 Chi-squared (χ^2) values for testing for homogeneity of hybrid necrosis rates within the TR or NTR groups in crosses with Cer10. 113

Table 4.8 The Chi-squared (χ^2) values from tests where the necrosis data from each TR in turn is omitted. 114

Table 4.9 The Chi-squared (χ^2) values from tests where the necrosis data from each NTR in turn is omitted. 115

Table 4.10 The Chi-squared (χ^2) values from tests comparing the degree of hybrid necrosis of each TR, and their group average, with 25_E01. 116

Table 4.11 The Chi-squared (χ^2) values from chi-squared tests comparing the degree of hybrid necrosis of each NTR, and their group average, with 25_E11. 116

Table 4.12 Summary of the average hybrid necrosis rates from crosses between Cer10 and recombinants. The data is shown both pooled for the TR and NTR groups and also with the data from TR 25_E11 and NTR 25_E01 omitted from their respective groups. 117

Table 4.13 The identity and position of the four informative markers associated with the copper tolerance locus and their genotypes in offspring of the D x A cross. 118

Table 4.14 The marker genotypes and recombination breakpoints in the recombinants (between marker e242 and the T locus) that were crossed with Cer10 to establish the degree of hybrid necrosis caused. Informative plants are highlighted in bold. 120

Chapter 5

Table 5.1 A summary of Dobzhansky-Muller incompatible genes whose biological functions have been identified. 128

Table 5.2 The fragment sizes (genotypes) in the parent plant lines A and D of the fifteen markers genotyped in the first round of marker screening. 135

Table 5.3 The fragment sizes (genotypes) in the parent plant lines A and D of the seven markers genotyped in the second round of marker screening. 136

Table 5.4 The genotypes of the three informative markers that are linked to copper tolerance. 138

Table 5.5 The genotype (nucleotide) at different bases of the three SNP markers screened for variation in plants A and D and selected offspring. 140

Table 5.6 The marker genotypes in parent plants A and D and in TRs and NTRs with regards to their associated tolerant/non-tolerant phenotype thereby revealing the recombination breakpoints in the D x A plants recombinant between e242 and T. 142

Chapter 6

Table 6.1 Identified pairs or sets of complementary genes whose diverged alleles interact in a Dobzhansky-Muller-type incompatibility resulting in hybrid dysfunction. 155

Table 6.2 The genotypes and location of the polymorphic markers identified in Cer10 and Cer35. Those shown in bold print were identified as co-segregating with the hybrid necrosis phenotype. 166

Table 6.3 The hybrid necrosis rates of the F_2 plants that were selected for BSA and of those plants which should have been selected. 168

Table 6.4 The frequency of the genotypes in the BSA high- and low- necrosis F_2 groups of the twenty-five polymorphic markers in Cer10 and Cer35. 170

Table 6.5 The genotypes and one-way ANOVA p-values for the three markers identified for QTLs controlling hybrid necrosis in Cerig. 173

Table 6.6 The frequency of the nine possible genotypes from the combination of markers e217 and e416 in the 137 F_2 hybrids with genotypes at these two markers and the average degree of hybrid necrosis (%) associated with each genotype. 175

Table 6.7 The number of parental (P) and recombinant (R) chromosomes present in the Cerig F_2 population in the combined genotypes of markers e217 and e416. 175

Table 6.8 The chi-squared (χ^2) values from comparing the degree of hybrid necrosis associated with the three recombinant genotypes of the markers e217 and e416 with the degree associated with the nearest non-recombinant genotypes. 177

Table 6.9 The five combined genotypes of markers e416 and e279 present in the Cerig F_2 population and their associated necrosis rates. 182

Appendices

Table A4.1 Raw data for hybrid necrosis from crosses between Cerig 10 or Cerig 35 and the tolerant (TR) and non-tolerant (NTR) recombinants. 199

Table A6.1 Chi-squared values for testing for homogeneity between replicates trays from the same F_2 x Cop cross, replicate F_2 x Cop crosses and between all replicate trays each F_2. 202

Figures

Chapter 1

Figure 1.1 The Dobzhansky-Muller model of genic incompatibility. 34

Chapter 2

Figure 2.1 The main physical features of *M. guttatus* 66

Figure 2.2 The morphological differences between the study populations Cerig and Copperopolis. 67

Figure 2.3 Maps showing the location of Copperopolis, Calaveras County, California 68

Figure 2.4 Maps from showing the location of Cerig-y-Drudion, Conwy, North Wales 69

Figure 2.5 Glasshouse plant growth methods 71

Figure 2.6 Phenotyping for Copper Tolerance 75

Figure 2.7 Photographs of a stigma on an *M. guttatus* plant showing the closing stigma system. 76

Figure 2.8 Agarose gel electrophoresis 83

Chapter 3

Figure 3.1 The phenotype of hybrid necrosis in progeny from a cross between a Cerig individual and either a copper tolerant or non-tolerant plant. 91

Figure 3.2 Variation in the degree of hybrid necrosis in three trays containing offspring from crosses between a Cop individual and three different Cerig individuals. 95

Chapter 4

Figure 4.1 The mean degree of hybrid necrosis in control crosses between grouped tolerant recombinants or non-tolerant recombinants and either Cerig35, Cerig10 or Cop, and in testcrosses between Cerig10 and each individual recombinant. 109

Figure 4.2 A genetic and physical map showing the positions of the four T locus markers identified by research reported in this chapter. 119

Figure 4.3 A map of Chromosome 9 of the *M. guttatus* genome showing the location of the location of the copper tolerance locus (T) and the five markers identified in this research as linked to the T locus. 122

Figure 4.4 Map showing the positions and recombination frequencies/distance in cM between the T locus and the four linked markers and the NEC locus. Courtesy of Kevin Wright, Duke University. 124

Chapter 5

Figure 5.1 The positions along scaffold_84 of the Mimulus genome sequence of the nine mRNAs whose sequences were used to design the fifteen markers screened for variation in plants A (copper non-tolerant) and D (copper tolerant) in the first round of screening. 132

Figure 5.2 The positions along scaffold_84 of the Mimulus genome sequence of the five mRNA markers and eight microsatellite markers screened for variation in plants A and D in the second round of screening. 133

Figure 5.3 The positions along scaffold_84 of the Mimulus genome sequence of the informative copper-tolerance linked markers that identify the window containing the NEC locus. 140

Figure 5.4 The position on scaffold_84 of the Mimulus genome of known R genes, the T locus markers identified in Chapters 4 & 5 and the recombination window containing NEC. 143

Figure 5.5 A summary of the results of the mapping, fine mapping and bioinformatics in Chapters 4 & 5 showing the informative markers on scaffold_84 of the Mimulus genome that narrow down the region containing NEC and the three candidate genes within this 59kb window. 144

Chapter 6

Figure 6.1 A histogram showing the frequency distribution of the residuals for the proportion of hybrid necrosis in crosses between Cop and 142 individuals from a segregating F_2 population of Cerig. 161

Figure 6.2 A histogram showing the frequency distribution of the residuals from the arcsin-transformed proportion of hybrid necrosis in crosses between Cop and 142 individuals from a segregating F_2 population of Cerig. 162

Figure 6.3 The wide variation in the distribution of the degree (proportion) of hybrid necrosis in crosses between Cop and 142 individuals from a segregating F_2 population of Cerig. 164

Figure 6.4 A histogram showing the frequency of the distribution of the degree (proportion) of hybrid necrosis in crosses between Cop and 142 individuals from a segregating F_2 population of Cerig. 165

Figure 6.5 The relative positions on the 14 linkage groups of the *M. guttatus* genome of the seventy-three markers screened for polymorphism in Cer10 and Cer35. 167

Figure 6.6 Maps of Chromosome 9 of the *M. guttatus* genome showing the location of markers e217 and e416 along with the location of NEC and T loci as identified by Chapters 4 & 5. 174

Figure 6.7 The average degree of hybrid necrosis associated with each of the three genotypes of markers e217 and e416 in the segregating F_2 population of Cerig. 174

Figure 6.8 The average degree of hybrid necrosis associated with the three genotypes of marker e416 in the segregating F_2 Cerig population. 178

Figure 6.9 The distribution of the degree of hybrid necrosis associated with each individual in the segregating F_2 population with respect to their genotype at marker e416. 179

Figure 6.10 A map of chromosome 12 of the *M. guttatus* genome showing the location of marker e279. 180

Figure 6.11 The average degree of hybrid necrosis associated with the two genotypes of marker e279 present in the segregating F_2 population of Cerig. 181

Appendix 7

Figure A7.1 Maps of the Mimulus genome showing the size (cM) of the fourteen chromosomes and the locations of the MgSTS markers screened. 208

Author's Declaration

Aspects of this thesis are based on a collaborative project between Kevin Wright, a PhD student based at Duke University, North Carolina, and myself. Provided below is a summary of the individual contribution each of us to the research reported in this thesis. I certify that all writing and data analysis included in this thesis is my own work.

Work	Performed By
Chapter 3	
Growing plants	Deborah Lloyd
Phenotyping for copper tolerance	Deborah Lloyd
Tissue sampling	Deborah Lloyd
Identifying Cerigs to use in crosses	Deborah Lloyd
Conducting crosses	Deborah Lloyd
Phenotyping for hybrid necrosis	Deborah Lloyd
DNA Extraction	Kevin Wright
Genotyping to identify recombinants	Kevin Wright
Chapter 4	
Growing plants	Deborah Lloyd
Phenotyping for copper tolerance	Deborah Lloyd
Tissue sampling	Deborah Lloyd
Conducting crosses	Deborah Lloyd
Phenotyping for hybrid necrosis	Deborah Lloyd
DNA Extraction	Kevin Wright
Genotyping to identifying recombinants	Kevin Wright
Chapter 5	
Growing plants	Deborah Lloyd
Conducting crosses	Deborah Lloyd
Scoring for hybrid necrosis	Deborah Lloyd
Design of new primers	Deborah Lloyd
DNA extraction	Deborah Lloyd
Screening markers	Deborah Lloyd
Bioinformatics	Deborah Lloyd
Screening SNP markers	Kevin Wright

Work	Performed By
Chapter 6	
Growing plants	Deborah Lloyd
Conducting crosses	Deborah Lloyd
Phenotyping for hybrid necrosis	Deborah Lloyd
Tissue sampling	Deborah Lloyd
Designing primers	Deborah Lloyd
DNA extraction	Deborah Lloyd
Marker Genotyping in BSA	Deborah Lloyd
Marker Genotyping of linked markers in all F_2	Deborah Lloyd

Abbreviations

'A' plant	Plant 'A' used in crosses (bred by Professor Macnair, a copper non- tolerant plant with the SB genetic background)
A. thaliana	Arabidopsis thaliana
ANOVA	Analysis of Variance
Avr	Avirulence protein
bp	basepairs
BSA	Bulked Segregant Analysis
BSC	Biological Species Concept
Cerig/Cer	'Cerig-y-drudion'
Cf-	*Cladosporium fulvum* resistance gene
cM	Centimorgans
Cop	'Copperopolis'
'D' plant	Plant 'D' used in crosses (bred by Professor Macnair, a copper tolerant plant with the SB genetic background and a section of the Cop genome)
D. melanogaster	*Drosophila melanogaster*
D. simulans	*Drosophila simulans*
DM	Dobzhansky-Muller
DM1	*Dangerous Mix 1*
DM2	*Dangerous Mix 2*
EMBOSS	European Molecular Biology Open Software Suite
EST	Expressed Sequence Tag
F_1	First filial generation
F_2	Second filial generation
Hdb2	*Hybrid breakdown 2*
Hdb3	*Hybrid breakdown 3*
Hmr	*Hybrid male lethality*
HR	Hypersensitive Response
Hw (a/b/c/d/e/f/g/h)1	*Hybrid weakness gene (a* to *h)*
Hw (a/b/c/d/e/f/g/h)2	*Hybrid weakness gene (a* to *h)*
JGI	Joint Genome Institute
JmjC	Jumonji domain
kb	kilobases
KEGG	Kyoto Encyclopaedia of gene and genomes
KOG	Eukaryotic Orthologous Groups
L. esculentum	*Lycopersicon esculentum*

L. pimpinellifolium	*Lycopersicon pimpinellifoliu*
L. saligna	*Lactuca saligna*
LG	Linkage Group
Lhr	*Lethal hybrid rescue*
M. guttatus	*Mimulus guttatus*
Mhr	*Maternal hybrid rescue*
mRNA	messenger RNA
N. longiflora	*Nicotiana longiflora*
N. suaveolens	*Nicotiana suaveolens*
N. tabacum	*Nicotiana tabacum*
NBS-LRR	Nucleotide-binding Leucine-Rich-Repeat
NCBI	National Centre for Biotechnology Information
Ne1	*Necrosis gene 1*
Ne2	*Necrosis gene 2*
***NEC* (locus)**	*Hybrid necrosis locus*
NPC	Nuclear pore complex
NT	Copper non-tolerant
NTR	Copper non-tolerant recombinant
O. sativa	*Oryza sativa*
Ovd	*Overdrive*
P. syringae	*Pseudomonas syringae*
PANTHER	Protein Analysis Through Evolutionary Relationships
PCD	Programmed Cell Death
PCR	Polymerase Chain Reaction
PFAM	Protein Family Database
Prdm1	*PR domain containing 1*
QTL	Quantitative Trait Loci
***R* gene**	Disease Resistance Gene
RF	Recombination Frequency
RI	Reproductive Isolation
RIN4	RPM1 Interacting Protein 4
RPM1	Resistance to *Pseudomonas syringae PV Maculicola*
RPS2	Resistance to *Pseudomonas syringae* 2
S. cerevisiae	*Saccahromyces cerevisiae*
SA	Salicylic acid
SB	Stinson Beach
Sc_	Scaffold (of *Mimulus* genome sequence)

SGE	Selfish Genetic Element
SNP	Single Nucleotide Polymorphism
***T* (locus)**	Copper tolerance locus
T	Copper tolerant
TAIR	Arabidopsis Information Resource
TR	Copper tolerant recombinant
X. maculates	*Xiphophorus maculates*
Xmrk2	*Xiphophorus melanoma receptor tyrosine kinase*
Zhr	*Zygotic hybrid rescue*

Chapter 1

General Introduction

The origin of species is a central theme in biology, yet still relatively little is known about the process at the molecular level. Understanding how and why barriers to gene flow arise during speciation is undoubtedly a formidable challenge but elucidating the genetic basis of such barriers will enable a mechanistic understanding of the evolutionary forces that create and maintain biodiversity.

Although universally famous, Darwin's work "The Origin of Species" (Darwin, 1859) is well known to focus more on adaptation than on the rise of new species. It was not until the introduction of Mayr's Biological Species Concept, which defines species as groups of interbreeding natural populations that are reproductively isolated from other such groups (Mayr, 1942), that insights into the process of speciation began to be realised.

1.1 The BSC and other species concepts

From the wealth of books and journal papers that deal with the subject of species concepts, that is, the different approaches to identifying species (see Coyne and Orr, 2004, de Queiroz, 1998), it is clear that this is a controversial issue in the field of evolution, with much debate over how best to define what constitutes a species. Many diverse concepts have been put forward and used in research. Coyne & Orr in their book 'Speciation' (2004) and de Queiroz in his chapter in 'Endless Forms' (1998) review some of the many species concepts that exist (summarised in table 1.1).

Whilst some of these concepts bear similarities to one another, others differ radically. For example, the Biological Species Concept (BSC) and the Isolation Species Concept (ISC) both define species in terms of reproductive isolation, whereas the Ecological Species Concept (ESC) defines species in terms of adaptation and natural selection, whilst other concepts, such as the Recognition Species Concept, use fertilization systems as the major approach to classifying species.

Under the Biological Species Concept (BSC), the key distinguishing feature of a species is that it is prevented from merging with other species by the presence of reproductive isolating mechanisms which form barriers to gene flow. Reproductive isolation (RI) is the product of all

such isolating barriers and thus is a central factor in evolutionary biology because of its role in determining gene flow between populations (Coyne and Orr, 1998).

Table 1.1 A summary of the main Species Concepts
(from definitions and citations found in (Coyne and Orr, 2004) and (de Queiroz, 1998))

Species Concept	Species …	Based on…	Notes
Biological Species Concept	… are groups of actually or potentially interbreeding natural populations that are reproductively isolated from other such groups (Mayr 1995)	Interbreeding	Emphasise that species are populations of interbreeding organisms.
Isolation Species Concept	… are systems of populations; the gene exchange between these systems is limited or prevented in nature by a reproductive isolating mechanism or perhaps by a combination of several such mechanisms. (Dobzhansky 1970)	Interbreeding	Mayr's definition (above) and Dobzhansky's definition both fall into this category and this definition has been incorporated into Mayr's definition (Mallet, 2006).
Recognition Species Concept	… is that most inclusive population of individual biparental organisms which shares a common fertilization system. (Paterson 1985)	Genetic or phenotypic cohesion	Species defined by factors that hold populations together rather than by those that isolate them. Emphasis on unification of species by specific mate recognition systems shared by conspecific organisms rather than the separation of species from each other due to RI between heterospecific organisms.
Evolutionary Species Concept	… is a single lineage of ancestral descendent populations or organisms which maintains its identity from other such lineages and which has its own evolutionary tendencies and historical fate (Wiley 1978, modified from Simpson, 1961)	Evolutionary Cohesion	Emphasises the extension of species through time and attempts to accommodate the observations that some populations appear to maintain their distinctness despite interbreeding with other populations.
Ecological Species Concept	… is a lineage (or a closely related set of lineages) which occupies an adaptive zone minimally different from that of any other lineage in its range and which evolves separately from all lineages outside its range (Van Valen 1976)	Evolutionary Cohesion	Emphasises the importance of ecologically based natural selection in maintaining species.

Table 1.1 continued…

Species Concept	Species …	Based on…	Notes
Cohesion Species Concept	… is the most inclusive population of individuals having the potential for phenotypic cohesion through intrinsic cohesion mechanisms (Templeton 1989)	Genetic or phenotypic cohesion	Uses components of the evolutionary, ecological, isolation and recognition definitions and emphasises the mechanisms that maintain evolutionary lineages by promoting genetic relatedness and determining the boundaries of populations with respect to micro- and natural selection.
Phylogenetic Species Concept 1	… is an irreducible (basal) cluster of organisms that is diagnosably distinct from other such clusters, and within which there is a paternal pattern of ancestry and descent (Cracraft 1989)	Evolutionary History	Emphasis on diagnosability regardless of whether characters are apomorphic. Some advocates conclude that species are not necessarily monophyletic.
Phylogenetic Species Concept 2	… is the smallest (exclusive) monophyletic group of common ancestry (de Queiroz & Donoghue 1988)	Evolutionary History	Based on distinctions between monophyly (inferred by presence of derived or apomorphic characters) and paraphyly.
Phylogenetic Species Concept 3	… is a basal, exclusive group of organisms all of whose genes coalesce more recently with each other than with those of any organisms outside the group, and that contains no exclusive group within it (Shaw, 1998)	Evolutionary History	Emphasis on diagnosability, regardless of whether the derived characters are apomorphic.
Genotypic Cluster Species Concept	… is a (morphologically or genetically) distinguishable group of individuals that has few or no intermediates when in contact with other such clusters (Mallet 1995)	Genetic or phenotypic cohesion	Independent of theories concerning the origin and maintenance of species. Emphasis on evidence, especially genetics, used to recognise species.

Whilst most evolutionary biologists adhere to the BSC, defining species on the basis of reproductive isolation barriers does have its limitations, as discussed by Coyne & Orr (2004), and is not without its critics. Indeed, Patterson's (1985) Recognition Species Concept is based on the cohesive factors that hold populations together and argues most fervently against the inclusion of postzygotic isolating barriers in defining a species. However, postzygotic reproductive isolating barriers have been shown to be important in sympatric populations, for example between two species of *Mimulus* (Ramsey et al., 2003), and could be important in maintaining species distinctness in cases where allopatric species that diverged from a common ancestor regain physical contact.

Other species concepts attempt to overcome the possible limitations of the BSC, but, as Mayr (1963) points out, the application of any species concept can be difficult, especially as evolution is a gradual process and in nature many populations are only part way towards attaining species status and may have acquired only partial attributes of the three most characteristic properties of species – RI, ecological difference and morphological distinction. Although it is entirely possible that certain aspects of all these species concepts are correct and that species cannot be assigned a simple definition but instead need to be considered in the light of several concepts, for the purposes of studying the process of speciation it is necessary to have an appropriate definition of a species. In actual fact, despite the diverse range of concepts that have been proposed and used as parameters in research, each concept ultimately defines a species by reference either to a reproductive community (inter-breeding/fertilisation/reproductive isolation/gene exchange) or to degree of difference (morphology/phenotype) (Mayr, 2001) thereby bringing the range of concepts back to just two basic principles: reproductive isolation and phenotype.

As the two populations of the yellow monkey flower *Mimulus guttatus* studied in this thesis are geographically isolated, one being from California, U.S.A, and one from Wales, U.K., one limitation of the BSC that is particularly pertinent to this thesis is whether or not it can be applied to allopatric taxa, whereby populations are geographically isolated and so it is unknown whether they could coexist in sympatry without exchanging genes. However, factors such as the interbreeding of migrants with local populations or even artificial hybridizations performed in experimental studies can enable the use of the BSC even in allopatric speciation studies. Furthermore, this issue with biogeography is not limited to the BSC but applies to every concept that is not based on phylogenetics and so adopting a different species concept may not remove this possible limitation. In addition, as this thesis is concerned with the genetic basis of a reproductive isolating barrier found when the two populations are crossed in glasshouse experiments, the most applicable species concept is one that defines species with reference to isolating barriers and RI. As such, the BSC is the most applicable species concept in this thesis. Moreover, most studies on RI adhere to the BSC, including nearly all those cited in this thesis, meaning that its use here will facilitate direct comparison and reference to the other studies on RI.

1.1.1 Biogeography and Speciation

Given the geographical isolation of the study populations in this thesis, it is worth considering the highly contentious issue in speciation that is biogeography. Speciation is usually divided into three groups, shown in table 1.2, based on biogeography, and consequently on the freedom of/restrictions on populations to exchange genes. There are alternative methods of classifying

modes of speciation, but, as Coyne and Orr (2004) state, these other classifications are still ultimately based on biogeography. For instance, modes of speciation can be classified by the evolutionary forces acting on species and resulting in RI, but the strength and nature of these forces are directly affected by the environment/habitat and thus by biogeography.

Table 1.2 Definitions for the three modes of speciation based on biogeography as described in Coyne & Orr (2004)

Mode of speciation	**Definition**
Allopatric:	complete geographic isolation and no gene exchange:
- vicariant	RI evolves after the geographic range of a species splits into two or more large, isolated populations
- peripatric	RI evolves after an isolated habitat is colonised by a few individuals or a small population becomes geographically isolated
Parapatric	partial isolation and limited gene exchange
Sympatric	no isolation and free gene exchange

In allopatric speciation, any evolutionary force causing populations to diverge can eventually yield speciation (Slatkin 1985 as cited by Coyne and Orr, 2004). These forces are likely to vary greatly in different geographical areas and so produce disparate selection pressures that will act on populations and, unhindered by gene flow, cause them to diverge and potentially attain intrinsic RI that will prevent successful interbreeding should they regain physical contact. Allopatric speciation is supported by the Dobzhansky-Muller model, described in greater detail later, which illustrates how it permits the evolution of incompatible alleles at complementary genes thereby resulting in hybrid dysfunctions (Coyne and Orr, 2004). This is the model that is applied to the hybrid lethality seen in the *M. guttatus* cross studied in this thesis (Macnair and Christie, 1983).

The conditions allowing parapatric and sympatric speciation to occur are more restrictive, as the forces that cause populations to diverge must be stronger than the gene flow causing them to fuse (Slatkin 1985 as cited by Coyne and Orr, 2004). However, because new mutations that are sensitive to genetic background or habitat or that are tightly linked to genes causing RI can continue to accumulate, gene flow does not halt population divergence (Rieseberg, 2001). Despite this, Coyne & Orr's (2004) review suggests that whilst sympatric speciation is both plausible and observed in nature, it is not common, with the exception of polyploid speciation in plants.

Whatever the status of the biogeography, whether the populations are living in sympatry or allopatry, the crucial element in the process of speciation is the development of reproductive isolating barriers that result in RI between populations.

1.2 Reproductive Isolation

With the BSC in mind, speciation can be defined as the process by which two identical populations diverge genetically to the point at which their subsequent merger is not possible due to reproductive isolation brought about by the acquisition of reproductive isolating barriers (Coyne and Orr, 1998, Wu and Ting, 2004). As such, for speciation to occur, barriers to gene flow between interbreeding populations must arise that sufficiently restrict gene flow enough to overcome its homogenizing effect thereby causing the populations to become reproductively isolated from each other (Scopece et al., 2008). The process is only complete when RI prevents gene flow in sympatry (Coyne and Orr, 1998). Accordingly, the crucial element in the process of speciation is evolution of isolating barriers, described in table 1.3, that prevent gene flow. Knowledge of the genetic basis of such barriers is therefore of great importance to an understanding of the dynamics of speciation (Christie and Macnair, 1984).

Although it is possible for one strong isolating barrier to be responsible for complete RI, in nature, individual reproductive isolating barriers are rarely sufficient on their own to cause complete RI and it is more usual for RI to result from a cumulative effect of several isolating mechanisms (Coyne and Orr, 1998, Coyne and Orr, 2004, Lowry et al., 2008). For example, in crosses between *Drosophila simulans* and *D. melanogaster*, pre-mating isolation tends to prevent the occurrence of matings and even when matings are successful oviposition often fails (Sawamura et al., 1993a). In addition, any offspring that do occur are sterile and, depending on the direction of the cross, one or other of the sexes is lethal (Sturtevant, 1920). This shows how a combination of isolating barriers can give rise to RI and indicates that speciation is unlikely to be a consequence of a single isolating barrier (Widmer et al., 2009). As a result, a major aim in speciation biology is determining the relative contribution and importance of prezygotic versus postzygotic barriers in RI.

In plants, prezygotic barriers tend to appear to be stronger than postzygotic barriers because they have a greater impact on reducing gene flow (Widmer et al., 2009). However, the contribution of prezygotic barriers to the cessation of gene flow might be overestimated because although they may have a major effect on reducing gene flow they are often incomplete and even small amounts of gene flow between populations may be sufficient to prevent divergence. The residual gene flow that remains after the action of prezygotic barriers can only be prevented by the action of postzygotic barriers, therefore rendering these extrinsic and/or intrinsic barriers

essential for cessation of gene flow and thus complete RI. In addition, the way in which the relative contributions of pre- and post-zygotic barriers are usually calculated, based on Coyne & Orr's approach (1989, 1998), assumes that because in nature each barrier acts sequentially each barrier can prevent only the gene flow not prevented by earlier acting barriers. Consequently postzygotic barriers, which by their very nature act later than prezygotic barriers, will always have a smaller overall contribution as they have less gene flow to act upon. This may bias the apparent importance of the barriers during the evolution of RI.

Table 1.3 Reproductive Isolating Barriers (from Coyne and Orr, 2004: Table 1.2)

Isolating Barrier	**Description/Mode of Action**
Prezygotic barriers act before fertilisation	
Premating isolating barriers	
impede gene flow before transfer of sperm or pollen to members of other species.	
1. Behavioural Isolation	includes all differences that lead to a lack of cross-attraction between members of different species, preventing them from initiating courtship or copulation.
2. Ecological Isolation	based on differences in species ecology, i.e., barriers that are direct by-products of adaptation to the local environment
a) Habitat isolation	species have genetic or biological propensities to occupy different habitats when they occur in same general area, thus preventing or limiting gene exchange through spatial separation during the breeding season. This isolation can be caused by differential adaptation, differential preference, competition, or combinations of these factors.
b) Temporal isolation	Gene flow between sympatric taxa is impeded because they breed at different times
c) Pollinator isolation	Gene flow between angiosperm species is reduced by their differential interactions with pollinators. This can occur via pollination by different species, or by pollen transfer involving different body parts of a single pollinator species
3. Mechanical Isolation	Inhibition of normal copulation or pollination between two species due to incompatibility of their reproductive structures. This incompatibility can result from lack of mechanical fit between male and female genitalia (structural isolation) or the failure of heterospecific genitalia to provide proper stimulation for mating (tactile isolation)
4. Mating System Isolation	The evolution of partial or complete self-fertilisation (autogamy) or the asexual production of offspring (apomixis) that can result in the creation of a new taxon or set of lineages. Not an isolating barrier in the same sense as the others in this table.

Table 1.3 *continued…*

Isolating Barrier	**Description/Mode of Action**
Postmating isolating barriers	
act after sperm or pollen transfer but before fertilisation	
1. Copulatory behavioural isolation	behaviour of an individual during copulation is insufficient to allow normal fertilisation
2. Gametic isolation	transferred gametes cannot effect fertilisation
a) Non-competitive gametic isolation	intrinsic problems with transfer, storage, or fertilisation of heterospecific gametes in single fertilisation between members of different species
b) Competitive gametic isolation	heterospecific gametes are not properly transferred, stored, or used in fertilisation only when competing with conspecific gametes
Postzygotic isolating barriers act after fertilisation	
1. Extrinsic	depends on the environment, either biotic or abiotic
a) Ecological inviability	hybrids develop normally but suffer lower viability because they cannot find an appropriate ecological niche.
b) Behavioural sterility	hybrids have normal gametogenesis but are far less fertile than parental species because they cannot obtain mates. Most often, hybrids have intermediate phenotypes or courtship behaviours that make them unattractive.
2. Intrinsic	A developmental problem in hybrids that is relatively independent of the environment
a) Hybrid inviability	hybrids suffer developmental difficulties causing full or partial lethality
b) Hybrid sterility	
i) Physiological sterility	hybrids suffer problems in the development of the reproductive system or gametes
ii) Behavioural sterility	hybrids suffer neurological or physiological lesions that render them incapable of successful courtship

It would seem, then, that both pre- and post-zygotic barriers can be crucial for the acquisition of RI, the relative contribution of each varies with each species and the amount of gene flow each barrier reduces is not necessarily a direct indication of the importance of the barrier: if there were two barriers, one which reduced 99% of gene flow and one which reduced 1%, the latter

barrier would be just as important as the first despite the large difference in contribution as it is this final 1% that makes RI, and therefore speciation, complete.

Investigations into the relationship between RI and genetic distance (the allelic substitutions per locus that have occurred during the separate evolution of two populations or species) in animal species have shown that, in general, prezygotic isolation evolves faster than postzygotic isolation (Coyne and Orr, 1989). Whether this is so in plants is not known, but the strength of postzygotic barriers does increase with increasing genetic distance amongst taxa which may indicate slow evolution (Moyle and Graham, 2005, Scopece et al., 2008). However, rapid evolution of postzygotic barriers does appear to occur in some groups as there is considerable variation in the strength of these barriers, perhaps because the rate of trait evolution depends on its genetic architecture, with traits that are controlled by few major genes having the potential to evolve faster than traits under the control of numerous genes with minor effect (Widmer et al., 2009).

So, determining the precise molecular basis for each individual isolating barrier and obtaining a mechanistic understanding of the evolutionary forces that caused it to evolve will reveal much about how and why these isolating barriers arise during speciation (Lowry et al., 2008).

This thesis is concerned with the inviability of hybrid progeny from a cross between two populations of *M. guttatus* and accordingly focuses on an intrinsic postzygotic isolating barrier. Many species pairs exhibit intrinsic postzygotic isolation, which is independent of ecological situation and affects the development of the hybrid making it either fully or partially sterile or inviable (Coyne and Orr, 2004, Fishman and Willis, 2001). There are three types of genetic differences that can potentially cause the reduction of hybrid fitness (fertility and viability) in intrinsic postzygotic isolation: different ploidy levels (Burton and Husband, 2000, Bomblies, 2006), chromosomal rearrangements (Coyne and Orr, 1998, Rieseberg, 2001) and genic incompatibilities (Coyne and Orr, 2004). These mechanisms vary in importance depending on the system but all three have been shown to have a major role in plant speciation (Rieseberg and Willis, 2007). However, the focus of this thesis is on the role of genic incompatibilities, as the crossing barrier between the study plant populations arises through this mechanism (Macnair and Christie, 1983).

Barriers to gene flow evolve gradually with genetic distance (Scopece et al., 2008). However, the very evolution of intrinsic postzygotic isolation posed a problem for early evolutionary biologists because natural selection acts to increase individual fitness and so should not directly favour or select for maladaptive traits such as premature death or sterility of hybrid offspring (Darwin, 1859, Orr, 1995, Presgraves, 2007). How intrinsic postzygotic isolation can evolve is therefore problematic as hybrids would have to pass through an intermediate step in which they

are unfit and this would be opposed by natural selection. As such, in order to evolve, hybrid sterility and inviability cannot be directly selected for, in which case how can the factors causing hybrid breakdown be maintained and increase in frequency enough to permit the divergence of populations (Coyne and Orr, 2004)? A solution to this problem was found in the form of a genetic model, first proposed by Bateson in 1909 then later formalised by Dobzhansky (1936, 1937) and Muller (1942), which is now widely accepted and hereafter referred to as the Dobzhansky-Muller (DM) model.

1.3 The Dobzhansky-Muller Model of genic incompatibilities

The DM model posits that hybrid dysfunctions (sterility or inviability) arise from deleterious epistatic interactions between incompatible alleles of complementary genes that have diverged so that the alleles function well in their normal within-species genetic background but fail to interact properly in the hybrid genome (Coyne and Orr, 2004, Wu, 2001b, Orr et al., 2004). As RI therefore evolves as a by-product of divergence, alleles causing hybrid dysfunctions can accumulate as they will not be directly under selection (Fishman and Willis, 2001). This model now forms the basis of a large majority of research into the genetics of intrinsic postzygotic isolation (Presgraves, 2006, Coyne and Orr, 1998) and has led to the accumulation of much evidential material to support genic incompatibilities resulting in intrinsic postzygotic isolation (see section 1.5).

Although they may represent only a small fraction of the genome, these DM genes can create strong barriers to gene exchange between diverging populations (Ortiz-Barrientos and Kane, 2007). There is some contention as to whether RI applies to every locus in the genome, a ‘whole genome’ concept (Wu, 2001a, Wu and Ting, 2004) or only to particular loci, the ‘genic' view, where a limited number of genes prevent interbreeding (Mayr, 2001, Rieseberg, 2001). Whilst two diverging populations are only true species when they are fully reproductively isolated from each other and cannot produce any functional hybrids, gene-flow barriers themselves act on discrete loci meaning the unit of isolation is not the entire genome but rather the chromosomal regions that harbour isolation loci (Mayr, 2001, Rieseberg, 2001, Rieseberg and Carney, 1998). Thus, the whole genome can be isolated but not by genes on every chromosome. However, sequences in the whole genome will diverge independently as in the ‘whole-genome’ concept, but only when the speciation process has been completed and gene flow has essentially ceased (Bomblies and Weigel, 2007). Partial RI indicates that the isolating mechanism may be a genic stepping-stone on the route to speciation. So, it seems clear that the genic and whole genome concepts are inexorably linked - genic incompatibilities can result in genomes from two species being isolated, as fertile/viable hybrids cannot be formed.

Whether the result of genes on a few chromosomes or on every chromosome, the isolating barriers are the same and it is the inability to produce viable or fertile hybrids which renders a species reproductively isolated and thus makes two species distinct from each other. It is also worth taking into consideration that understanding speciation is not simply a matter of studying RI and finding the loci responsible for isolation but also requires the investigation of whether there exists stable genetic differentiation in the face of potential gene flow when species are in sympatry (Mallet, 2006).

The simplest model of DM genic incompatibilities is based on the interaction between two complementary genes (Figure 1.1) in which an ancestral species with *aa* at one locus and *bb* at another splits into two genetically-identical isolated populations. This could result from a number of factors such as geographical isolation, where the populations have become physically isolated from one another, or ecological isolation, where the populations are separated by moving into different habitats or niches that cause them to coincide very rarely (Coyne and Orr, 2004). As the two populations are now isolated, they will experience different selection pressures according to their different and distinct environments and so new mutations will be differentially selected for and fixed in each population. In addition, mutations are random and so it is unlikely the same mutation would occur in two isolated populations (Coyne and Orr, 2004). Consequently, each population will acquire a new range of unique alleles that the ancestral species did not possess.

For instance, a new mutation might occur in one population causing an *A* allele to appear and become fixed with the resultant *Aabb* and *AAbb* genotypes being viable and fertile. A second mutation in either population causes a *B* allele to arise and become fixed. In figure 1.1(i), this mutation arises in the second population, and the resultant *aaBb* and *aaBB* genotypes are also viable and fertile (Coyne and Orr, 1998, Wu and Ting, 2004). Although these new *A* and *B* alleles are viable in the genetic background of the population in which they arose they have not occurred together and so, upon secondary contact between the two populations, an *AaBb* hybrid is formed which, if the *A* and *B* alleles are incompatible, could be partially or completely inviable or sterile. Alternatively, as described in figure 1.1(ii), one population may retain the ancestral genotype *aabb* whilst both alleles in the second population experience mutations and become *AAbb* and then *AABB*, both of which are perfectly fit. However, in this situation the *B* allele has never before been in contact with the ancestral *a* allele, and these alleles may be incompatible when brought together in the *AaBb* hybrids (Orr, 1995). So, the new *A* and *B* alleles are viable in the genetic background of the population in which they arose but may be incompatible with an allele they have not co-existed with in their ancestral species genetic background.

Thus, intrinsic postzygotic isolation between two populations results from epistatic interactions between functionally diverged incompatible alleles of complementary genes (Presgraves, 2007), and so for genic incompatibilities to arise there must be at least two independent substitution events in either the same or different lineages (populations) and at least one locus from each lineage must be involved (Coyne and Orr, 2004, Dobzhansky, 1937, Muller, 1942). Although the DM model describes incompatibilities between fixed genes, it is not strictly necessary for these genes to become fixed in order for them to cause hybrid dysfunctions. Indeed, polymorphism at DM genes has been found in several cases including *Crepis* (Hollingshead, 1930), *Mimulus* (Macnair and Christie, 1983), *Nicotiana* (Bomblies and Weigel, 2007) and *Gossypium* (Stelly, 1990).

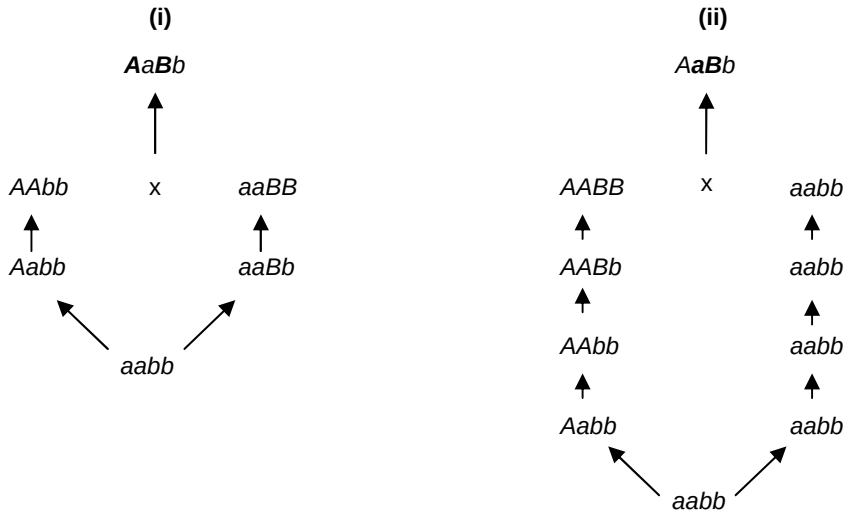


Figure 1.1 The Dobzhansky-Muller model of genic incompatibility.

Incompatible alleles (shown in bold) causing hybrid dysfunctions can arise in either (i) both populations where a random mutation at the *A* locus in population 1 and the *B* locus in population 2 gives rise to new alleles that are independently fixed in each lineage and hybrid incompatibility is caused by interactions between the *A* and *B* alleles OR (ii) one population where population 2 retains the ancestral genotype whilst two random mutations give rise to new alleles in population 1 so that the *B* allele, which does not arise until after the *A* allele is fixed, has never occurred with the ancestral *a* allele until hybridisation whereupon the interaction between *B* and *a* causes hybrid dysfunction. [NB when discussing DM genic incompatibilities, uppercase letters do not indicate that an allele is dominant but that it is derived i.e. has undergone mutation and is consequently different from the ancestral allele represented by lowercase letters]

In the DM model the terms “dominant” and “recessive” are applied to describe how the effects of the incompatibility between alleles *A* and *B* are modified by the compatible alleles *a* and *b*

(Turelli and Orr, 1995). The incompatibility is said to be dominant if any genotype including at least one *A* and one *B* allele expresses the same level of dysfunction. Conversely, the incompatibility is said to be recessive if the presence of at least one *a* or *b* allele is adequate to mask the incompatibility; that is, only *AABB* suffers hybrid dysfunction. Partially recessive/dominant incompatibilities are those in which the extent of the dysfunction is partially reduced by *a* and *b*, so that the double heterozygote *AaBb* has higher fitness than *AABB*, and genotypes that are homozygous-heterozygous (*aABB* and *AABb*) are intermediate (Fitzpatrick, 2008, Turelli and Orr, 1995).

The central assumption of the DM model is that alleles cause no sterility or inviability when in the normal species genetic background but lower fitness when brought together with alleles from another species (Orr, 1995). The development of postzygotic isolation by such genic incompatibilities shows several patterns as described by Orr (1995).

First, all hybrid incompatibilities must be asymmetric because the required fertility/viability of all intermediate steps constrains which incompatibilities are possible and the ancestral and derived alleles at one locus cannot be incompatible with alleles at the same locus (Muller, 1942, Orr, 1995). This means there must always be at least two loci involved in the interaction.

Second, each individual genic incompatibility does not necessarily have a devastating effect on the hybrid, but may only cause a slight reduction in hybrid viability/fertility, so that full postzygotic isolation results from the accumulation of several incompatibilities (Coyne and Orr, 2004). As such, the degree of postzygotic RI, as well as the number of genes involved, should "snowball" and increase faster than linearly with time because any new substitution in one species is potentially incompatible with the alleles at every locus that has previously diverged in the other species. Hybrid incompatibilities will accumulate even faster if the interaction involves more than two loci. This has important consequences for genetic analysis as it could mean the number of genes required to cause strong RI could be overestimated (Coyne and Orr, 1998, Orr, 1995).

Third, although the DM model usually describes two-locus interactions, in fact these genic incompatibilities can be complex interactions involving more than two genes and a hybrid might need to have particular alleles at three or more loci before its fertility/viability is affected so that the hybrid genotype *Abc* might be sterile/inviable while any genotype consisting of any other combination of these alleles might be perfectly fertile/viable. Indeed, complex interactions among three or more genes should be common and there is growing evidence supporting complex incompatibilities from work in *Drosophila* (Coyne and Orr, 1998).

1.4 Forces driving the evolution of hybrid incompatibilities: selection, drift and genomic conflict

The DM model requires only that substitutions occur and predicts neither the nature of the genetic changes nor the evolutionary mechanisms that lead to hybrid incompatibilities and assumes nothing about whether they are brought about by natural selection or drift (Orr, 1995).

1.4.1 Genetic drift

If the new forms of the genes have no effect on fitness, incompatibilities could accumulate simply by random genetic fluctuations. Genetic drift is the random variation in gene frequency from one generation to another in a population and could potentially play a role in speciation if a small group of individuals becomes isolated from the main population. When combined, the random variation in gene frequency, the higher rate of drift in the small group (due to the small number of individuals) and the prevention of gene exchange between the small group and main population cause the small group to genetically drift from the rest of the species. Over time, the group could potentially drift enough to become sufficiently different from the main population to be reproductively isolated.

However, there is considerable evidence from a wide range of studies (reviewed by Coyne and Orr, 2004) which suggests that the role of genetic drift in speciation is rare and that selection plays a much more important role.

1.4.2 Natural Selection

Unlike with genetic drift, in this instance the new forms of the genes are beneficial and thus divergence is a result of positive natural selection. Indeed, the evolutionary analysis of the genes identified thus far as contributing to hybrid incompatibilities strongly indicates that speciation results from positive natural selection (Orr et al., 2004). For example, both *Hmr* and *Nup96* from *Drosophila* have been shown to be under recurrent positive selection which could have led to functional divergence in multiple species (Barbash et al., 2004, Barbash et al., 2003, Maheshwari et al., 2008).

It is generally thought that natural selection is unlikely to select directly for postzygotic isolation in sympatry as this would entail selecting for a maladaptive trait (death/sterility of offspring), although a model by Macnair (1987a) demonstrates that it is possible. Instead, the classic view of the evolution of hybrid incompatibilities is that natural selection acts indirectly so that

reproductive isolating barriers arise as a by-product of selection for an adaptive character (Coyne and Orr, 2004).

Alternatively, natural selection acting on an adaptive character may cause linked genes to also be selected for due to hitch-hiking - the process by which genes can increase in frequency and spread through a gene pool if they are linked to a beneficial allele whose frequency is increasing through natural selection (Macnair, 1987b, Fay and Wu, 2000). Thus, proximity on a chromosome may allow neutral, or even deleterious, genes to be selected for due to an advantageous gene nearby. This phenomenon has been implicated in the evolution of the hybrid inviability studied in this thesis (see Macnair and Christie, 1983) as described in greater detail later. Likewise, analysis of quantitative trait loci (QTL) potentially involved in RI has found that QTLs for different traits cluster in particular areas of the genome providing evidence for linkage among QTLs for different traits. Selection on one of these traits can therefore induce selection on the other traits, which has the potential to lead to divergence in characters that are not the direct target of selection (Widmer et al., 2009).

One way in which this selection-driven speciation process might proceed is through the evolution of ecotypes - genetically different forms of populations that are adapted to a particular set of environmental conditions - and their ultimate transition to new species (Abbott and Comes, 2007). A well documented example is the evolution of heavy metal tolerance in plants growing on mine spoil heaps (Antonovics et al., 1971, Macnair, 1997).

Notwithstanding the normally toxic levels of metals such as zinc, lead and copper on mine tailings, there are usually plants found growing there. The essential difference between species able to evolve tolerance and those that do not is the possession of genetic variance for tolerance, meaning tolerant individuals exist even in a non-tolerant population and thus will be selected for in a metal contaminated soil (Macnair, 1987b). Ecotypes frequently occur in habitats that are geographically isolated from one another and so ecotypic variation is not eroded by gene flow, but ecotypes can also exist in sympatry where there is more likelihood of interbreeding and gene flow (Abbott and Comes, 2007). In both cases natural selection will be important in establishing divergence and RI.

So, natural selection can cause rapid adaptive divergence of populations by selecting for adaptive traits that result indirectly in RI and is now regarded to be of central importance to the process of speciation (Abbott and Comes, 2007).

However, this view of hybrid incompatibilities arising as by-products of selection on an adaptive trait is being challenged by accumulating evidence, albeit suggestive rather than conclusive, that genomic conflict can drive divergence and cause hybrid incompatibilities.

1.4.3 Genomic Conflict

In genomes of individuals there exist so-called selfish genetic elements (SGEs) such as transposable elements or repetitive DNAs that over-replicate or other selfish elements, termed segregation distorters, that subvert Mendel's law of segregation by preferentially entering gametes during meiosis. Genomic conflict arises because these genes manipulate host reproduction to facilitate their own transmission to the next generation at the expense of other genes, or even the organism as a whole, so other host genes often evolve to suppress the selfish DNAs or compensate for their deleterious effects (Burt and Trivers, 2006, Presgraves, 2010). It has been suggested that this "arms race" between SGEs and suppressors can lead to the evolution of hybrid dysfunctions either because there are incompatibilities between host genes that have evolved to silence or mitigate the selfish genes or because SGEs are not suppressed in the genome of the hybrid (Burt and Trivers, 2006, McDermott and Norr, 2010, Presgraves, 2010). This latter mechanism, of which there are three examples in *Drosophila* (see Presgraves, 2010 for a review), is particularly interesting as it proposes an alternative mechanism to the current classical view of hybrid dysfunction arising from genes that have diverged and whose new alleles are incompatible.

The role of genomic conflict in known cases of hybrid incompatibilities is reviewed by Johnson (2010) and Presgraves (2010) and indicates some cases are more likely than others to have arisen by genomic conflict. There is evidence to suggest genomic conflict has caused hybrid dysfunctions to arise in plants (Barr and Fishman, 2010, Case and Willis, 2008), *Drosophila* (Phadnis and Orr, 2009, Presgraves, 2010) and mice (Mihola et al., 2009, Oliver et al., 2009). If genetic conflict is the force driving the evolution of hybrid dysfunctions, then the first steps in the evolution of hybrid sterility or inviability may be nearly neutral or even deleterious rather than adaptive (Presgraves, 2010).

Clearly there is evidence to support the roles of natural selection and genomic conflict in the evolution of hybrid dysfunction and further opportunities to study the effects of these forces can be gained by identifying the genetic elements involved in hybrid sterility and inviability.

1.5 Examples of hybrid incompatibilities and DM genes

There are many examples of hybrid incompatibility, but until recently the genetics of speciation and the actual genes involved in intrinsic postzygotic RI were unknown. As recently as 2006, no pair of DM genes had been reported (Brideau et al., 2006). However, the availability of genome sequences has aided evolutionary biologists/geneticists in identifying and characterising several genes that cause hybrid sterility and inviability, summarised in table 1.4, and genomic studies

are revealing more about the complexities of the incompatible genes involved in postzygotic isolation. Despite this, still relatively little is known about how 'speciation' genes cause RI, but genes that have so far been characterised have proposed modes of action based on their function in the normal, within-species genetic background. Hybrid incompatibilities can arise as a result of direct interactions of incompatible gene products or from the mis-regulation of genes as a result of incompatibilities between regulatory sequences and factors that control gene expression (Bomblies, 2006). As such, the complementary genes involved in DM interactions are ordinary genes, although those that have been identified have been found to belong to an array of functional groups and may not necessarily have a large phenotypic effect. Such genes need not have drastic effects on hybrid fitness as any particular incompatibility might lower hybrid fitness by only a small amount (Coyne and Orr, 1998) and total RI is the result of the additive effects of several individual incompatibilities. How these genes cause hybrid dysfunctions is best described through examples of such genes identified by studies. Although a number of genes causing hybrid sterility have been identified (for some examples see table 1.4), this thesis focuses on those DM genes that have been identified as contributing to hybrid lethality, and particularly hybrid necrosis.

1.5.1 Hybrid lethality genes in Drosophila

Much of the research into genic incompatibilities was, until recently, carried out in *Drosophila* where post-mating RI is ensured by both the inviability and sterility of hybrids.

Two rescue mutations, *Hybrid male rescue* (*Hmr*) which is X-linked (Barbash et al., 2004, 2000, Hutter and Ashburner, 1987, Hutter et al., 1990, Orr and Irving, 2000) and autosomal *Lethal hybrid rescue* (*Lhr*) (Brideau et al., 2006, Nolte et al., 2008, Prigent et al., 2009, Takamura and Watanabe, 1980) were found to restore viability in normally lethal male hybrids from a cross between *D. melanogaster* females and *D. simulans* males. Both are recessive, loss-of-function mutations indicating that a single incompatibility between the wild-type alleles of these genes means that Lhr^{+}, which has functionally diverged in *D. simulans*, and Hmr^{+}, which has functionally diverged in *D. melanogaster*, interact deleteriously in hybrid males to cause lethality (Barbash et al., 2000, Brideau et al., 2006, Hutter et al., 1990, Prigent et al., 2009).

As the affected males have a pathology that resembles that of mitotic mutants in *D. melanogaster* it is thought that their death results from a mitotic defect either due to cell cycle arrest or due to a failure in chromosome condensation during mitosis (Bolkan et al., 2007, Orr et al., 1997, Prigent et al., 2009). This would concur with the apparent role of the *Hmr+* and *Lhr+*

Table 1.4 Examples of genes identified as contributing to Dobzhansky-Muller incompatibility interactions leading to hybrid dysfunctions. Where known, interacting gene pairs/sets are given. Available information from the literature on the identity or function of the genes and the evolutionary force leading to the divergence of these genes between the species involved is also given.

Species 1	**Gene 1**	**Identity/ Function**	**Species 2**	**Gene 2**	**Identity/ Function**	**Hybrid Phenotype**	**Notes**	**Refs.**
A. thaliana Columbia accession (Col)	*HPA* At5g10330	Histidine biosynthesis	*A. thaliana* Cape Verde Island accession (Cvi)	*HPA2* At5g71920	Histidine biosynthesis	Lethality	Divergent evolution of duplicate genes	Bikard (2009)
D. simulans	*Nup96*	Nucleoporin (component of NUP107 subcomplex)	*D. melanogaster*	Unknown factor(s) on X chromosome	Unknown	Lethality	Byproduct of adaptive evolution driving divergence	Presgraves et al. (2003); Tang & Presgraves (2009)
D. simulans	*Nup160*	Nucleoporin (component of NUP107 subcomplex)	*D. melanogaster*	Unknown factor(s) on X-chromosome	Unknown	Lethality	Byproduct of divergent coevolution with *Nup96*	Tang & Presgraves (2009)
India rice varieties	*Hwa1*	Unknown	India rice varieties	*Hwa2*	Unknown	Lethality		Oka (1956)
X. maculates	*Xmrk2*	Tyrosine kinase	*X. helleri*	*Regulatory gene (R)*	Tumour suppressor	Lethality		Wittbrodt et al. (1989); Coyne & Orr (2004); Schartl (2008)
D. simulans	*Maternal hybrid rescue (mhr)*	Possible suppressor of *Zhr* meiotic drive	*D. melanogaster*	*Zygotic hybrid rescue (Zhr)*	Satellite DNA involved in maintaining hetero-chromatic states (probable role in mitosis)	Lethality in F_1 females	Possible genomic conflict	Sawamura et al. (1993a, 1993c); Presgraves (2010)

Table 1.4 *continued*

Species 1	Gene 1	Identity/ Function	Species 2	Gene 2	Identity/ Function	Hybrid Phenotype	Notes	Refs.
D. simulans	*Lethal hybrid rescue (Lhr)*	Involved in maintaining hetero-chromatic states	*D. melanogaster*	*Hybrid male rescue (Hmr)*	DNA binding	Lethality in F_1 males	Strong evidence of positive selection	Watanabe (1979); Takamura & Watanabe (1980); Nolte et al. (2008); Prigent et al. (2009); Hutter & Ashburner (1987); Hutter et al. (1990); Orr et al. (1997); Orr & Irving (2000); Barbash et al. (2000, 2003, 2004); Bolkan et al. (2007); Maheshwari et al. (2008); Brideau et al. (2006); Sawamura (2000)
G. hirsutum L. (D or A genome-carrying *Gossypium* species)	*Le_1 or Le_2*	Unknown	*G. davidsonii* & *G. klotzschianum* (D3 genome carrying)	*Le_2^{dav}*	Unknown	Lethality/ necrosis		Lee (1981); Stelly (1990); Samora et al. (1994)
A. thaliana strain Uk-3	*DM1*	*R* gene	*A. thaliana* strain Uk-1	*DM2*	Unknown, possible cellular partner for *DM1*	Necrosis	Involvement of plant immune system	Bomblies et al. (2007)
G. hirsutum	*Le_3*	Unknown	*G. barbadense* cv Coastland R4-4	*Le_4*	Unknown	Necrosis		Song et al. (2009)
L. saligna	*Rin4*	*R* gene guardee (interacts with *R* proteins)	*L. sativa*	Unknown	Unknown	Necrosis	Involvement of plant immune system	Jeuken et al. (2009)

Table 1.4 *continued*

Species 1	Gene 1	Identity/ Function	Species 2	Gene 2	Identity/ Function	Hybrid Phenotype	Notes	Refs.
L. pimpinellifolium	*Cf-2*	*R* gene	*L. esculentum*	*RCR3*	Encodes a cysteine endoprotease, probable guardee of *Cf-2* R protein	Necrosis	Involvement of plant immune system	Kruger et al. (2002)
Lycopersicon	*Cf-9*	*R* gene	*Lycopersicon*	Unknown	Unknown	Necrosis	Involvement of plant immune system	Wulff et al. (2004)
T. aestivum L. varieties	*Ne1*	Unknown	*T. aestivum* L. varieties	*Ne2*	Unknown	Necrosis		Hermsen (1963); Chu et al. (2006); Singh et al. (2000); Jiang, Q. et al. (2008)
India/Japanese Rice varieties	*hwb1*	Unknown	India/ Japanese Rice varieties	*hwb2*	Unknown	Necrosis/ Weakness (breakdown)		Oka (1956)
O. glumaepatula	*hwf1*	Unknown	*O. sativa* L. cv. Taichung 65	Unknown	Unknown	Necrosis/ Weakness (breakdown)		Sobrizal & Yoshimura (2009)
O. nivara	*hbd1*	Unknown	*O. sativa* L. ssp *japonica* Koshihikari; ssp *indica* cultivar Nona Bokra	Unknown	Unknown	Necrosis/ Weakness (breakdown)		Matsubara et al. (2006) ; Miura et al. (2008)

Table 1.4 *continued*

Species 1	Gene 1	Identity/ Function	Species 2	Gene 2	Identity/ Function	Hybrid Phenotype	Notes	Refs.
O. sativa Assam accession	*hwg1*	Unknown	*O. sativa* spp. *japonica* Sasanishiki cultivar	*hwg2*	Unknown	Necrosis/ Weakness (breakdown)		Fukuoka et al. (2005)
O. sativa Thai cultivar Col. No 15	*hwd1*	Unknown	*O. sativa* Nepalese cultivar Siborunauli 1	*hwd2*	Unknown	Necrosis/ Weakness (breakdown)		Fukuoka et al. (1998)
O. sativa L. ssp *indica* Habataki cultivar	*hbd2*	Casein kinase 1	*O. sativa* spp. *japonica* Koshihikari & Sasanishiki cultivars	*hbd3*	*R* gene (NBS-LRR)	Necrosis/ Weakness (breakdown)	Possible involvement of plant immune system	Matsubara et al. (2007); Yamamoto et al. (2007); Yamamoto et al. (2010)
O. sativa L. ssp *indica* IR24	*Hwe1*	Possible rice blast resistance gene	*O. sativa* L. ssp *japonica* variety Asominori	*hwe2*	Unknown	Necrosis/ Weakness (breakdown)	Possible involvement of plant immune system	Kubo & Yoshimura (2002)
O. sativa L. ssp *indica* Milyang 23	*hwh1*	Glucose-methanol-choline oxido-reductase	*O. sativa* L. ssp *japonica* Tong 88-7	*hwh2*	Hexose transporter	Necrosis/ Weakness (breakdown)	Involvement of plant immune system	Jiang, W. et al. (2008)
O. sativa cv Jamaica	*Hwc1*	Unknown	*O. sativa* cv Norin 8, Taichung 65, Nipponebare	*Hwc2*	Unknown	Necrosis/ Weakness		Sato & Morishima (1987); Ichitani et al. (2001, 2007); Saito et al. (2007)

Table 1.4 *continued*

Species 1	Gene 1	Identity/ Function	Species 2	Gene 2	Identity/ Function	Hybrid Phenotype	Notes	Refs.
O. sativa L. ssp *indica*	*S5-i*	Aspartic protease	*O. sativa* L. ssp *japonica*	*S5-j*	Aspartic protease	Sterility	Triallelic system at one locus	Chen et al. (2008)
N. intermedia	*dfe*[d]	Unknown	*N. crassa*	*dma*[d]	Unknown	Sterility	Byproduct of adaptive evolution	Dettman et al. (2008, 2009)
S. bayanus	*AEP2*	Mitochondrial translation protein	*S. cerevisiae*	*OLI1*	ATP synthase	Sterility (breakdown)	Byproduct of adaptive evolution	Lee et al. (2008)
S. cerevisiae	*MRS1*	Splices mitochondrial COX1 introns	*S. paradoxus*	Mitochondrial gene(s)	Unknown	Sterility (breakdown)	Cytonuclear incompatibility	Chou et al. (2010)
S. cerevisiae	*AIM22*	Ligase required for mitochondrial protein lipolyation	*S. paradoxus*/ *S. bayanus*	Mitochondrial gene(s)	Unknown	Sterility (breakdown)	Cytonuclear incompatibility	Chou et al. (2010)
O. sativa L. ssp *indica* variety IR24	*hsa1*	Unknown	*O. sativa* L. ssp *japonica* variety Asominori	*hsa2 & hsa3*	Unknown	Female Sterility		Kubo & Yoshimura (2005)
D. mauritiana	*Odysseus (OdsH)*	Homeobox domain chromosome condensation	*D. simulans*	Y-chromosome repetitive hetero-chromatic DNA	Unknown	Male sterility	Possible genomic conflict	Ting et al. (2004, 1998); Bayes & Malik (2009)

Table 1.4 *continued*

Species 1	Gene 1	Identity/ Function	Species 2	Gene 2	Identity/ Function	Hybrid Phenotype	Notes	Refs.
M. guttatus 'IM62'	*nad6*	Unknown	*M. nasutus* 'SF'	*restorer-of-male-fertility (rmf)*	Unknown	Male sterility	Cytoplasmic male sterility	Fishman & Willis (2006); Case & Willis (2008)
M. guttatus	*hybrid male sterility 1 (hms1)*	Unknown	*M. nasutus*	*hybrid male sterility 2 (hms2)*	Unknown	Male sterility		Sweigart et al. (2006)
Mus m. musculus	*Prdm1*	Histone-3- lysine-4 methyl-transferase or meiotic driver suppressor	*Mus m. domesticus*	Unknown	Unknown	Male sterility	Possible genomic conflict	Mihola et al. (2009); Oliver et al, (2009)
O. sativa spp. *indica*	*SaM+SaF+*	F-box protein	*O. sativa* L. ssp *japonica*	*SaM-*	Ubiquitin-like E3 ligase-like protein	Male sterility	Triallelic system at one locus	Long et al. (2008)
D. pseudoobscura bogota (Bogota)	*Overdrive (Ovd)*	DNA-binding	*D. pseudoobscura pseudoobscura* (USA)	Unknown	Unknown	Male sterility and segregation distortion	Genomic conflict	Phadnis & Orr (2009)
T. californicus populations	*Cytochrome C (CYC)*	Electron transport system protein	*T. californicus* populations	*Cytochrome C oxidase (COX)*	Electron transport system protein	Breakdown in multiple fitness traits	Cytonuclear incompatibility	Harrison & Burton (2006)

gene products: HMR is a DNA-binding protein related to a family of MYB-related DNA-binding transcriptional regulators and LHR is involved in maintaining heterochromatic states.

It is possible that *Hmr* and *Lhr* may be the only major-effect genes and their lethal effect requires a hybrid genetic background in which altered chromosome morphology and chromatin structure due to the cumulative effects of species-specific differences results in hybrid male lethality. Alternatively, this apparently simple two-gene interaction may in fact be more complex and there is some evidence to suggest this is so. For example, the involvement of more than one gene from *D. melanogaster* is indicated by the apparent contribution to the suppression of male hybrid lethality by a locus on chromosome 1(AB)-1n, which could be an unrelated gene or could be another allele of *Hmr* (Barbash et al., 2000, Hutter et al., 1990) and that deletion of *Hmr*+, although rescuing the lethality of hybrid males, does not return cell cycle progression completely back to normal (Bolkan et al., 2007). Furthermore, Watanabe (1979) found that the lethality of hybrids between these species was independently affected by different genetic factors in addition to the *Lhr* gene. These were the previously discovered Sex Ratio (*SR*) factor and daughterless (*da*) factor (Watanabe and Yamada, 1977). Thus, hybrid lethality in this cross may be enhanced by a multilocus interaction involving additional genes.

The reciprocal cross between *D. melanogaster* and *D. simulans*, where the latter is the female parent, results in lethality in female hybrids which is caused by genes other than *Hmr* and *Lhr*. Two further hybrid rescue mutations, *maternal hybrid rescue* (*mhr*) in *D. simulans* and *zygotic hybrid rescue* (*zhr*) in *D. melanogaster* both rescue the female hybrids that die as embryos (Sawamura et al., 1993a, 1993c). Again, these are both recessive mutations and hybrid embryonic lethality in the female hybrids results from incompatibility between *D. melanogaster* X-chromosome and the maternal effect of *D. simulans* due to the wild type alleles at these loci (Sawamura et al., 1993c). The pathology of the lethal embryos is typical of maternal/zygotic transition failure (Sawamura, 2000, Sawamura et al., 1993b). The *Zhr*+ gene has been localised to a region rich in satellite DNA and is deleted in the rescue mutation i.e. a loss-of-function mutation (Sawamura, 2000). There is strong evidence to suggest that the *Zhr* locus directly causes hybrid lethality through failure of a satellite DNA to form or maintain a normal heterochromatic state resulting in mitotic/chromosomal segregation failure in early precellularised embryos (Ferree and Barbash, 2009).

Interestingly, it also seems that the maternal cytoplasm from *D. simulans* lacks factors that are compatible with and necessary for the proper segregation of the *D. melanogaster* X-linked satellite DNA. This means that non-coding heterochromatic DNA can directly cause hybrid incompatibilities and thereby contribute to speciation (Ferree and Barbash, 2009). However, as this is not a protein-coding gene but a block of satellite DNA repeats and is a likely example of the involvement of a selfish gene from *D. melanogaster* causing hybrid incompatibility as the

D. simulans cytotype has not evolved the necessary regulators to suppress this selfish DNA (Presgraves, 2010), it is questionable as to whether this can be considered a DM gene.

It is now widely accepted that postzygotic isolation in this cross between *Drosophila* species is caused by wild-type alleles at hybrid rescue genes and that one set of loci (*mhr* from *D.simulans* and *zhr* from *D. melanogaster*) cause hybrid embryonic lethality whilst another set (*Hmr* from *D. melanogaster* and *Lhr* from *D. simulans*) cause hybrid larval lethality (Orr and Irving, 2000). This reveals that RI can rely on a small number of genes. However, these rescue mutations, with the exception of *Hmr* (Barbash and Ashburner, 2003), do not restore the fertility of the viable hybrids that are produced in either direction of the cross.

These rescue mutations imply a simple developmental basis for RI, because if RI depends on many independent developmental problems or incompatibilities the chance of recovering a single mutation that can simultaneously correct all of these problems seems remote. A simple developmental basis suggests a simple genetic basis. However, it is interesting to consider that rescue mutations need not necessarily be alleles of speciation genes but could be site-suppressors – loci that, when mutated, can override the effects of the genes causing inviability, for example by allowing the use of alternative metabolic pathways (Orr and Irving, 2000). The accumulation of variation at many sets of genes in different taxa makes it a reasonable assumption that RI is irreversible, so what is particularly fascinating about these rescue mutations is that they are in effect reversing reproductive isolation, and causing evolution to retrace its steps, by restoring the fitness of hybrids (Orr and Irving, 2000). This suggests that evolution is not a unidirectional process but possesses plasticity and, in the event of certain mutations, can seemingly reverse, taking an evolutionary step backwards (Watanabe, 1979). However, it is not known how frequent these rescue mutations might be. They could be widespread but not yet identified due to lack of sufficient genetic resources in other species or they could be rare. Indeed, the *Hmr* mutation was only found to occur in natural populations at a rate of 5% (Hutter and Ashburner, 1987).

Several additional *Drosophila* genes have now been identified as being involved in hybrid dysfunctions. For instance, Presgraves et al. (2003) and Tang & Presgraves (2009) identified two autosomal genes, *Nup96* and *Nup160* respectively, in *D. simulans* that are incompatible with one or more genes on the X-chromosome of *D. melanogaster*, and result in hybrid lethality. Both these genes are viability-essential within their species, encoding a nuclear pore complex (NPC) protein (a nucleoporin) and directly interact with each other as part of the NUP107 subcomplex, a subset of interacting nucleoporins that together form a stable component of the NPC which mediates molecular traffic between the nucleus and the cytoplasm (Tang and Presgraves, 2009). As they interact directly with each other, it is possible that Nup96 and Nup160 are incompatible with the same X-linked factor(s) from *D. melanogaster*. It is not known

whether this is an example of two distinct two-locus hybrid incompatibilities or whether they are components of a single more complex incompatibility, although theory predicts that complex hybrid incompatibilities should evolve more readily that simple ones (Orr, 1995). Either way, it would seem that an incidental by-product of adaptive divergent coevolution among interacting components of the *Drosophila* NPC means the *D. simulans* Nup96, and possibly Nup160, is no longer compatible with the unknown interacting factor(s) encoded by the *D. melanogaster* X chromosome (Presgraves et al., 2003). This work also reinforces the view that the primary function of 'speciation' genes is a normal biological function and not simply one of preventing hybridisation and that natural selection is the driving force in their evolution.

However, Presgraves et al. (2003) used the *Lhr* rescue mutation to restore the viability of hybrid males in crosses between *D. melanogaster* females and *D. simulans* males in order to investigate these other potential hybrid incompatibilities and, although the authors identified twenty small regions of the *D. simulans* autosomal genome capable of causing complete hybrid inviability and have gone on to provide fine-scale genetic, molecular and evolutionary analysis of the first of these genes (*Nup96*), it is questionable as to how relevant these genes are to speciation if *Lhr/Hmr* alone, along with possible modifiers, are sufficient to cause complete inviability as only those substitutions that occur prior to RI are truly relevant to speciation. Then again, rescue mutations in nature that result in viable hybrids could restore gene flow between *Drosophila* species which may render these other hybrid incompatibilities crucial in preventing this gene flow although even this role is questionable as the viable hybrids that result from rescue mutations remain sterile.

1.5.2 Hybrid lethality genes in plants

Although much of the research into DM incompatibilities has been performed in *Drosophila*, studies in plants are increasingly yielding a number of valuable insights into the genetic complexity of the interactions leading to hybrid dysfunctions.

In plants, hybrid failure due to inviability or sterility can be attributed to both simple and complex genetic interactions (see the review by Lowry et al., 2008) and among many species pairs RI does not result from a single mechanism but is a consequence of a large number of pre- and post-zygotic barriers (Widmer et al., 2009). Recent research has revealed many insights into the genetic basis of such barriers, with a significant contribution arising from studies involving the model plant genus *Mimulus*.

1.5.2.1 Mimulus- a model system for evolutionary studies

The flowering plant genus *Mimulus* was first proposed as a potential model species at a conference in 1947 (Vickery, 1978) and now possesses a rich history of ecological and evolutionary genetic research due to species within the genus having been used in a range of studies on ecology, inheritance, population genetics and metal tolerance (see Wu et al., 2007 for a review).

As described by Vickery (1978), the *Mimulus* genus comprises around 150 species and is split into ten to twelve sections each of which contains one or more complexes of closely related species which in turn contain numerous ecologically and morphologically distinct but interfertile species making them well suited for studying ecological and evolutionary genetics in nature. Possessing wide phenotypic variation, species in this genus occupy a diverse range of habitats from aquatic to desert to alpine, can be annual, herbaceous perennial or woody perennial and have a range of mating systems from completely outcrossing to selfing to asexual (Wu et al., 2007). The *Mimulus guttatus* species complex represents much of this diversity.

Part of the *Simiolus* section, which consists of six species groups and shows the most speciation by aneuploidy and/or polyploidy of all the sections in the *Mimulus* genus (Vickery, 1995), the *Mimulus guttatus* species complex contains a group of closely related species whose centre of diversity is California but which are now also naturalised in Eastern United States, Britain, mainland Europe and New Zealand. Members of this complex vary in ploidy and are highly polymorphic, varying both within and between populations in features such as height and leaf and flower morphology, and occur in a range of habitats from the seashore through to mountainous regions at elevations of 3000 metres and appear to be colonizing plants often invading disturbed habitats and new areas such as the toxic tailings of copper mines (Vickery, 1978).

The most common species in the *Mimulus guttatus* complex is *M. guttatus* Fischer ex DC (Vickery, 1978). The original description of this species by Fischer in 1812 and De Candolle in 1813 was of plants collected in the Aleutian Islands and it is not inconceivable that the first escapes emanated from plants derived from this original collection, either accidently or deliberately for use as garden ornamentals, and rapidly became naturalised and widespread (Tokarska-Guzik and Dajdok, 2007).

M. guttatus is the presumed progenitor of other more geographically restricted members of the complex including *M. nudatus*, *M. glaucescens*, *M. laciniatus*, *M. nasutus*, *M. platycalyx* and *M. cupriphilus* (Macnair et al., 1989). Whilst *M. guttatus, M. nudatus*, and *M. glaucescens* are primarily outbreeders, *M. laciniatus*, *M. nasutus*, *M. platycalyx* and *M. cupriphilus* are primarily

inbreeders (Macnair et al., 1989). Reproductive isolation between these species is largely brought about by their breeding system, with the exception of *M. nudatus* which is isolated from other members of the complex by postzygotic incompatibilities (Macnair, 1992). However, *M. guttatus* sometimes develops partial postzygotic reproductive isolation between certain populations that can, on occasions, be as strong as the barriers between *M. guttatus* and other species in the complex (Vickery, 1964, Vickery, 1978). Indeed, Vickery's (1978) examination of crossing barriers between species in the *Mimulus* genus revealed two main results. First, crossing barriers were usually observed in interspecific hybridisations but were often only partial in their effect on hybrid fitness. Second, the degree of postzygotic isolation varied amongst populations of the same species and geographically dispersed populations of the same species could have greater or lesser intrinsic postzygotic RI than interspecific crosses. Thus, crossing barriers typically only partially reduce hybrid fitness and show variation within species (Wu et al., 2007).

Due to its diversity, relatedness to other model systems and its ease of propagation making it a suitable study plant, *M. guttatus* is a leading model system for studying ecological and evolutionary genetics and was selected by the Joint Genome Institute (JGI) for whole-genome sequencing (http://genome.jgi-psf.org/mimulus/mimulus.home.html).

1.5.2.2 Copper tolerance and hybrid lethality in M. guttatus

The first direct genetic analysis of hybrid incompatibilities in *Mimulus* was performed by Macnair & Christie (1983) upon the discovery of a partial crossing barrier between a copper tolerant Californian population 'Copperopolis' and a copper non-tolerant Welsh population 'Cerig-y-Drudion' during a study on the genetic basis of copper tolerance in California populations of *M. guttatus* (Macnair, 1983). This crossing barrier and the link between copper tolerance and hybrid lethality form the basis of the research in this thesis.

The evolution of copper tolerance in M. guttatus

The phenomenon of heavy metal tolerance in plants, such as the well known example of copper tolerance in *M. guttatus*, has been extensively studied (Antonovics et al., 1971) and demonstrates the power of natural selection in adapting plant populations to specific environmental factors.

The ability to evolve metal tolerance is due to the presence of genetic variance for tolerance, even in non-tolerant populations. Copper tolerant populations of *M. guttatus* have been found

growing at a number of abandoned copper mines in California (Allen and Sheppard, 1971) due to the presence of a single major gene for copper tolerance, although tolerance is also affected by other genes (modifiers) which are hypostatic to the copper tolerance gene (Macnair, 1983). All populations on soil contaminated with copper have one-hundred per cent tolerant plants but there is usually a steep cline at mine boundaries so that the frequency of tolerant individuals in off-mine environments is low (Macnair et al., 1993; Macnair, 1987b).

Since the number of tolerant individuals in a non-tolerant population is small, the initial colonization of the copper mines likely involved a small number of individuals. However, the new population would have been able to expand rapidly as the mine environment is novel and, as very few other species have evolved copper tolerance, largely uncolonized meaning interspecific competition is low (Macnair, 1987b).

In addition, the mine population was, and is, subject to strong natural selection. Allen & Sheppard (1971) showed that there exists very powerful selection for copper tolerance on copper soil at the germination and seedling stage which would have lead to the rapid fixation of genes conferring tolerance. Many genes loosely linked to the tolerance gene would also have been taken to fixation by a process known as hitch-hiking (Macnair, 1987b).

Thus, the evolution of copper tolerance in *M. guttatus* was possible due to its possession of genetic variation for this character and was brought about by natural selection.

However, tolerance to a particular metal might not be enough to enable a species to survive on a mine as there are other factors associated with the soil conditions, such as low nutrient and water availability or potentially even the presence of elevated levels of other heavy metals. If the edaphic conditions are such that a species cannot grow, mutations in genes controlling tolerance to these other factors will be necessary before that species can colonise the mine (Macnair, 1987b). So, as well as copper tolerance, there exists selection for other characters that improve the adaptedness of the plant to the mine environment.

M. guttatus is the probable progenitor of the endemic species *M. cupriphilus*, a copper tolerant species restricted to two small copper mines in Calaveras County California, located at 10.9 km from Copperopolis (Macnair et al., 1989). An obligate annual that is smaller, highly branched, and more floriferous than *M. guttatus*, *M. cupriphilus* flowers earlier than local populations of *M. guttatus*, a trait that is dominant and probably a way of escaping drought conditions later in the season (Macnair and Gardner, 1998). When the fitness of these species is compared under optimal conditions, *M. guttatus* is much fitter, that is it grows larger and sets more seed. However, these two species only coexist in an environment that is not optimal for *M. guttatus*. Plant size is determined by the microenvironment and in these conditions (dry soil, normally

toxic levels of copper, poor nutrient level) all plants are small and produce a relatively small number of seeds. When this microenvironment renders plant size less than nine centimetres, *M. cupriphilus* is fitter than *M. guttatus*. As the environment improves and plants are able to grow larger, *M. guttatus* is fitter than *M. cupriphilus* because, as an adaptation to the non-optimal mine conditions, *M. cupriphilus* has developed an early flowering strategy in which seed production by small plants has been optimised at expense of vegetative growth that would enable greater seed production at bigger size. This demonstrates how adaptation and selection play a role in speciation (Macnair et al., 1989).

The Copperopolis mine in Calaveras County, California, has been colonised by a large *M. guttatus* population which has formed the copper tolerant ecotype 'Copperopolis' (Allen and Sheppard, 1971, Macnair et al., 1993). It is interesting to note that, as with *M. cupriphilus*, the copper tolerant ecotype Copperopolis population also flowers earlier than local non-tolerant *M. guttatus* populations, potentially causing temporal isolation between the populations. This implies that the ecotype is also under selection for traits that enable best survival and reproduction in the mine environment which exerts very different pressures from those experienced by the non-mine non-tolerant populations of *M. guttatus*, showing how differing environments can lead to the fixation of different genes. However, between the ecotypes and endemics there is little difference in the degree of adaptation to the chemical environment of the soil and the availability of water seems to be the most important factor controlling plant fitness (Macnair and Gardner, 1998).

Copper tolerance and hybrid lethality

In the cross between the copper tolerant (T) Californian population Copperopolis (Cop) and the non-tolerant (NT) Welsh population Cerig originally studied by Macnair & Christie (1983), the F_1 hybrids affected by the partial crossing barrier were found to grow normally until the four-leaf stage at which point the leaves began to go yellow and became progressively necrotic with further growth inhibited. Plants affected early nearly always died whilst those affected later were able to survive for longer and in the absence of competition some grew very slowly for many months. This phenotype of the affected hybrids is very similar to a common type of postzygotic barrier in plants called hybrid necrosis, or sometimes referred to as hybrid weakness, which has characteristic phenotypes including cell death, necrosis, wilting, chlorosis (yellowing), and reduced growth rate, and often results in lethality meaning few necrotic hybrids reach reproductive maturity (Bomblies and Weigel, 2007).

Macnair & Christie (1983) found that this crossing barrier only appeared in crosses between copper tolerant (T) and non-tolerant (NT) plants and there was a genetic linkage between the

tolerance gene and the necrosis phenotype. Traits that are controlled by only a few major genes have the potential to evolve faster than traits under the control of numerous genes with minor effects (Widmer et al., 2009) and indeed copper tolerance is governed by a single major gene (Macnair, 1983). Crossing experiments between tolerant and non-tolerant plants revealed that the hybrid necrosis was caused by DM interaction(s) between alleles at the copper tolerance gene, or a gene tightly linked to it, from Cop interacting with unknown loci from the Cerig population. As such, hybrid necrosis could either be a pleiotropic effect of the tolerance gene itself or the effect of a tightly-linked gene that became fixed in the Cop population via hitch-hiking when the tolerance locus was under positive selection. As Wu et al. (2007) state, revealing whether the copper tolerance locus is the same gene or merely tightly linked to the hybrid necrosis gene will help determine the role of natural selection in establishing this barrier.

The degree of hybrid necrosis, that is the proportion of total hybrid offspring that show symptoms of hybrid necrosis, was found to vary depending on the Cerig parent, indicating there is a degree of polymorphism in the Cerig populations. In some crosses with Cop the hybrid necrosis is almost 100% meaning that gene flow is almost entirely eliminated and consequently RI is almost total between Cop and Cerig whilst in other crosses necrosis is almost zero, meaning genes can flow freely between the populations and RI is almost non-existent (Macnair and Christie, 1983; personal observation). It is interesting to see a similarity here with crosses conducted between two *Nicotiana* species where hybrid lethality was found to vary due to one individual being heterozygous at a single locus (Bomblies and Weigel, 2007) and also evidence of polymorphism at the complementary *Le* genes that cause necrosis in cotton (Stelly, 1990). Clearly this occurrence of partial RI due to allelic polymorphism in one population is not unusual.

Due to this polymorphism within the Cerig necrosis genes, this intrinsic postzygotic isolating barrier can only partially prevent gene flow between these two populations and so may represent an early-stage reproductive isolation barrier in the evolution of RI between these two *M. guttatus* populations.

Macnair & Christie (1983) proposed a simple genetic model, which conforms to the DM model, for how hybrid necrosis in this cross could arise from a genic incompatibility:

The original tolerant parent, Cop, is homozygous for the tolerance gene (*T*) which is linked to another gene (*S1*) which produces a synthetic lethal (i.e. hybrid necrosis) when in combination with another unlinked gene (*A*). The Cop genotype is therefore $T^{S1}T^{S1}aa$. The Cerig parent is non-tolerant homozygous, *tt*, and must be heterozygous for *A* as not all the F_1 die. The Cerig genotype is therefore *ttAa*. When Cop and Cerig are crossed, the $T^{S1}tAa$ hybrids die because *S1* and *A* are brought together, whilst the $T^{S1}taa$ are fully viable as there is no *A* allele present.

A key point from Macnair and Christie's (1983) work is that it again highlights the potential role of selection, showing how RI and speciation may be initiated easily, even in the face of considerable gene flow in sympatry, if an adaptive gene, or a gene tightly linked to it, results in the development of an isolating barrier. In this instance, it could be that strong selection for copper tolerance in the mine population (Cop) has indirectly caused the evolution of this isolating barrier due to hitch-hiking of the necrosis gene, a direct effect of its tight linkage with the tolerance locus.

Christie and Macnair (1984) identified a second independent simple genetic system that also gave partial RI. Here, a single gene in Copperopolis plants was found to interact with a gene from a Utah population to cause seedling death, via stunted growth and necrosis, before the first true leaves had expanded, and both populations were polymorphic at these genes. Christie and Macnair (1987) went on to study the distribution and degree of polymorphism of the genes from both these systems and found that both the distribution of component genes and genetic polymorphism was geographically widespread.

As copper non-tolerant populations of *M. guttatus* cannot inhabit the copper soils that tolerant plants grow in, these populations have a degree of prezygotic isolation through ecological (habitat) isolation. The effectiveness of this habitat disparity as a barrier to gene flow depends on the degree of cross pollination and consequently on whether the two populations have the same flowering time/ pollinator/ flower morphology. It has been found that Copperopolis flowers earlier than local, non-mine populations, so this may act as a prezygotic barrier (temporal isolation). However, this does not mean that the partial intrinsic postzygotic barrier of hybrid necrosis is not an important contribution in the development of total RI between these populations.

Clearly, the Cop and Cerig populations are isolated geographically and so it is not known to what extent other isolating barriers might have developed. Nevertheless, it is clear that genetic differentiation is having an impact on gene flow when these populations are artificially brought together. Moreover, the Cerig genes were also found in another population, Napoleon, which is only 10km from the Copperopolis population (Christie and Macnair, 1987) meaning that the interacting genes may be in the same geographical area as the necrosis gene from the Copperopolis population.

1.5.2.3 Hybrid necrosis in other plant systems

Hybrid necrosis has been found in a number of other systems where it may affect the F_1 or, in a phenomenon termed hybrid breakdown, not appear until the F_2. Although very few genes identified as causing necrosis have been characterised, studies have nonetheless revealed

some insights into hybrid necrosis and the phenotypic similarity in hybrid necrosis cases across a wide range of plant species suggests a common underlying mechanism (Bomblies et al., 2007).

Studies on hybrid necrosis have showed that it can be genetically simple; in fact, where the genetic architecture has been explored, most of the cases of hybrid necrosis are due to two-locus DM incompatibility interactions. A classic example is the between-species hybrid necrosis in the F_1 hybrids between *Crepis tectorum* and *C. capillaris* (Hollingshead, 1930) but there are many other examples including in wheat (Hermsen, 1963), *Nicotiana* (Masuda, 2007, Yamada et al., 1999), rice (Fukuoka et al., 1998, 2005, Ichitani et al., 2001, 2007, Jiang W et al., 2008, Kubo and Yoshimura, 2002, Matsubara et al., 2006, Miura et al., 2008, Oka, 1956, Sobrizal and Yoshimura, 2009, Yamamoto et al., 2007), lettuce (Jeuken et al., 2009), *Arabidopsis* (Alcazar et al., 2009, Bomblies et al., 2007), barley (Wiebe, 1934) and *Gossypium* (Song et al., 2009). Furthermore, the substitution within an allele can in itself be simple. For example, in rice the gain of deleterious activity of an allele causing hybrid dysfunction can be caused by the change of a single amino acid from isoleucine in one cultivar to lysine in another (Yamamoto et al., 2010) and by alleles that differ in just two nucleotides (Chen et al., 2008). This effect of a simple substitution is not specific to hybrid necrosis genes and is found in other systems. For example, the hybrid sterility gene *MRS1* from *S. cerevisiae* encodes a gene product required for splicing specific introns in the mitochondrial Cox1 protein, and it is thought that the incompatibility is due to a change in the splicing specificity of the MRS1 protein, caused by three nonsynonymous mutations in the amino acid sequence (Chou et al., 2010). It would also seem that the location of hybrid necrosis genes in the genome is diverse, meaning there are no restrictions on which part of the genome can be involved. In rice species alone, for example, genes for hybrid weakness have been identified on eight different chromosomes.

However, even simple genetic interactions can lead to complex physiological effects. For example, hybrid necrosis in wheat is caused by the interaction of two dominant complementary genes, Ne_1 and Ne_2 (Hermsen, 1963) yet the protein expression profile of hybrid necrosis revealed a complex cellular network covering a broad range of metabolic processes including protein transport, folding and assembly, antioxidation, photosynthesis, protein biosynthesis, carbohydrate metabolism, signal transduction and DNA/RNA modification with proteins being both up- and down-regulated (Jiang Q et al., 2008). This shows how the interaction of just two incompatible alleles can affect the proteome in many ways and result in hybrid necrosis.

The incompatible alleles causing necrosis can be recessive or dominant. Recessive alleles typically cause hybrid breakdown (necrosis in the F_2 generation or later) as in *Phaseolus vulgaris* (see Bomblies and Weigel, 2007, reference number 50) and rice (Fukuoka et al., 1998, 2005, Jiang W et al., 2008, Kubo and Yoshimura, 2002, Matsubara et al., 2006, Miura et al.,

2008, Oka, 1956, Sobrizal and Yoshimura, 2009) whilst dominant alleles cause necrosis in the F_1 generation as in *Crepis* (Hollingshead, 1930), *Gossypium* (Song et al., 2009, Stelly, 1990), and wheat (Hermsen, 1963).

The action of dominant alleles gives rise to the phenomenon of gene dosage whereby the incompatibility phenotype increases in proportion to the number of incompatible alleles present in the hybrid genome. This is a source of great variation in the extent of the necrosis phenotype in several crosses. For example, in rice, hybrid weakness in crosses between two cultivars, Jamaica and Norin 8, is caused by two semi-dominant complementary genes: *Hwc1* from Jamaica and *Hwc2* from Norin 8. The necrosis phenotype ranges from weak to severe depending on the alleles at these loci with the severe phenotype only being found when *Hwc1* was homozygous; when this gene was heterozygous the severity of the necrosis depended on whether *Hwc2* was homozygous or heterozygous. This shows that *Hwc1* contributed more strongly to the severity of hybrid symptoms than *Hwc2* (Ichitani et al., 2007). In wheat, hybrid necrosis varies from severe, including seedling lethality, to weak where there may be just some leaf yellowing or even no symptoms at all with a plant remaining phenotypically nearly normal throughout its development in spite of its genotype of necrosis (Hermsen, 1963). This is dependent on the dosage of the dominant necrosis alleles, which could total two ($Ne_1ne_1Ne_2ne_2$), three ($Ne_1Ne_1Ne_2ne_2$ and $Ne_1ne_1Ne_2Ne_2$), or four ($Ne_1Ne_1Ne_2Ne_2$), and the more dominant necrosis alleles present in the genotype, the greater the degree of necrosis (Hermsen, 1963). Similarly, in lettuce four levels of necrosis were found depending on the alleles at the two interacting loci (Jeuken et al., 2009).

However, there are additional factors that contribute to the severity of the necrosis phenotype. In wheat, for example, the severity of necrosis between the two extremes of the phenotype is almost continuous (Hermsen, 1963). The large variation in the severity of necrosis is largely due to the multiple alleles of the two genes. So far, three Ne_1 alleles (Ne_1^w, Ne_1^m *and* Ne_1^s) and five Ne_2 alleles (Ne_2^w, Ne_2^{mw}, Ne_2^m, Ne_2^{ms}, Ne_2^s) have been found (Chu et al., 2006, Hermsen, 1963) that differ in expression causing weak (*w*) medium (*m*) or strong (*s*) necrosis. This phenomenon, know as allelic expressivity, combined with a gene dosage effect can cause the phenotypic separation of classes according to degree of necrosis to become less apparent and gives rise to a more continuous distribution of necrosis (Hermsen, 1963).

Additional complexity to a simple genetic architecture can be found through the heterozygous nature of some DM genes. For example, in crosses between *Nicotiana sanderae* and *N. longiflora* the progeny are usually all lethal yet with one individual *N. longiflora* plant half the progeny were healthy. This suggests this individual was heterozygous at a single locus with one allele being incompatible with a locus in *N. sanderae* resulting in lethality (Bomblies and Weigel, 2007). As previously described, Macnair & Christie (1983) also found evidence of polymorphism

in the DM genes that interact to cause necrosis in crosses between two populations of *Mimulus guttatus*. However, in this cross the genetics are more complex in that a small number of genes are involved in one of the populations. Further evidence of polymorphism at complementary genes is found in the *Le* alleles that cause necrosis in *Gossypium* (Stelly, 1990).

Another factor affecting the severity of the necrosis phenotype is the time of onset of the first necrosis symptoms, which is variable although the cause of this is as yet unknown. For instance, Song et al (2009) found that in *Gossypium* crosses, the earlier the symptoms of lethality were manifested, the more severe they became. Variation in the time of onset has been found in other systems (Hermsen, 1963, Macnair and Christie, 1983) and as hybrid necrosis is often progressive (Hermsen, 1963) it is intuitive that the earlier the symptoms appear the more time they have to develop and so cause a more severe phenotype.

The degree of hybrid necrosis can be influenced by the environment. Many studies of hybrid necrosis have found that it is temperature sensitive with the phenotype being alleviated by higher temperatures in several genera including *Gossypium* (Phillips, 1977, Song et al., 2009, Stelly, 1990), lettuce (Jeuken et al., 2009), *Arabidopsis* (Alcazar et al., 2009, Bomblies et al., 2007), rice (Ichitani et al., 2007) and *Nicotiana* (Yamada and Marubashi, 2003, Yamada et al., 1999). However, the temperature at which the necrosis phenotype is suppressed appears to vary according to the genera. For instance, in the cross between the rice cultivars Jamaica and Norin 8, hybrids recover from weakness at a threshold temperature of 29-30^{o}C (Saito et al., 2007) whilst necrotic symptoms in hybrids from crosses between *Arabidopsis* species are alleviated at lower temperatures of 20-23^{o}C (Alcazar et al., 2009, Bomblies et al., 2007). The effect of temperature on hybrid incompatibility is not limited to hybrid necrosis, or indeed to plants. For example, the rescue of inviability by *Hmr* is temperature sensitive with male rescue being virtually complete at 18^{o}C but severely depressed at temperatures above 25^{o}C, at which point even viable hybrid females become inviable and die during metamorphosis (Hutter and Ashburner, 1987, Hutter et al., 1990). Interestingly this is the opposite of the effect of temperature in hybrid necrosis as in this case the low temperature is associated with hybrid rescue.

1.5.2.4 Hybrid necrosis and the plant immune system

The physiological mechanism for suppression of hybrid necrosis under high temperatures could be explained by another fascinating link between many of these studies: the involvement of the plant immune system. It is well known that disease resistance is heat sensitive, with the plant immune system being suppressed at higher temperatures rendering the plants more susceptible to pathogen infection (Zhu et al., 2010). Thus, if hybrid necrosis is caused by aberrant activation of the immune response it could explain why the phenotype is alleviated at higher temperatures.

The strong similarities in appearance between hybrid necrosis and plants responding to pathogen infection (Bomblies and Weigel, 2007) were observed as long ago as 1930, and led to the suggestion that the plant immune system was involved in hybrid failure (Kostoff, 1930) and in recent years there has been a resurgence of this theory. There is now a growing body of evidence, both biochemical and molecular, to support the theory that hybrid necrosis is a plant version of autoimmunity and results from the inappropriate activation of the plant immune system in the absence of pathogens due to genetic incompatibilities in the hybrid genome (Bomblies, 2009, Bomblies and Weigel, 2007, Ispolatov and Doebeli, 2009). Hybrid necrosis arising from aberrant interactions involving plant immune genes has been found in tomato (Kruger et al., 2002, Wulff et al., 2004), *Arabidopsis* (Alcazar et al., 2009, Bomblies et al., 2007), *Phaseolus* (Hannah et al., 2007), *Nicotiana* (Masuda, 2007), lettuce (Jeuken and Lindhout, 2002, Jeuken et al., 2009) and rice (Yamamoto et al., 2010).

When infected with a pathogen, a plant can mount a rapid resistance response due to the presence of disease resistance (*R*) genes in the host plant genome which perceive a pathogen-derived effector molecule which is specific to an individual pathogen and encoded by an avirulence (*Avr*) gene in the pathogen genome. This 'gene-for-gene' interaction elicits the hypersensitive response (HR), a form of programmed cell death (PCD) of infected, and sometimes neighbouring, cells, which causes the rapid collapse of tissue in order to restrict the ability of pathogens to grow and spread through living tissue (Greenberg, 1997, Dangl et al., 1996). The majority of *R* genes in plants reside in clusters and encode nucleotide-binding leucine-rich repeat (NBS-LRR) proteins (McHale et al., 2006) and have been shown to be the most likely temperature-sensitive component in defence responses (Zhu et al., 2010).

Although the activation of *R* gene products by pathogen effectors is well documented (Dangl and Jones, 2001, Jones and Dangl, 2006), there is little evidence for a direct interaction between *R* gene products and the *Avr* gene products, which gave rise to the formation of the 'guard hypothesis' (van der Biezen and Jones, 1998). This model proposes that the R proteins interact, or guard, another protein, the guardee, which is the target of *Avr* protein binding. There are two mechanistic scenarios, described by Dangl & Jones (2001), for the guard hypothesis. Firstly, the R protein could bind its guardee constitutively but disengage upon Avr binding to the guardee resulting in an active R protein. In this instance, the guardee would negatively regulate the R protein and the signalling pathways it mediates. Alternatively, binding of the Avr to the guardee could recruit the R protein to the Avr/target complex thereby triggering R protein activation. One known example of disease resistance progressing via a guard/guardee mechanism is resistance to *P. syringae* in *Arabidopsis* (Belkhadir et al., 2004). Here, the guardee protein RIN4 is associated with two R proteins, RPM1 and RPS2, and negatively regulates their inappropriate activation. The interactions of the *P. syringae* Avr proteins with RIN4 cause posttranslational modification, in the form of hyperphosphorylation, which in turn

activates the R proteins. Interestingly, pathogen effectors may have more than one target in the host cell, not all of which are associated with R proteins and this guard hypothesis means that multiple R proteins could associate with the same host protein complex, and hence with each other (Dangl and Jones, 2001).

However, plant pathogen effectors have now been shown to mimic plant transcription factors which, in susceptible plants, turn on genes that enable pathogen growth, but in resistant plants (i.e. those with the relevant *R* genes), turn on *R* genes. One example, reviewed by Greenshields & Jones (2008), reveals that the pathogen effector AvrBs3 targets a so-called *upa* box in the promoter of a cell-size regulator gene. However, in resistant plants this same *upa* box is found in the promoter of the *Bs3 R* gene so that AvrBs3 activates transcription of this gene triggering localised PCD. Similar examples have been found elsewhere suggesting this could be a common mode of action.

Thus, the plant defence response has evolved to be both sophisticated and complex with multiple mechanisms aimed at limiting pathogen success.

The process of hybrid necrosis bears some of the distinctive hallmarks of PCD. For example, wheat hybrids have been found to have increased levels of superoxide, a molecule associated with oxidative stress, PCD and the HR (Khanna-Chopra et al., 1998) whilst necrotic tobacco hybrids have been found to have chromatin condensation and DNA fragmentation (Marubashi et al., 1999). In addition, constitutive immune response activation can cause dwarfing, due to suppression of growth through promotion of cell death, and spontaneous lesion formation (autonecrosis) both of which are found in necrotic hybrids (Jones and Dangl, 2006). Although plants use PCD in normal development and during the HR during pathogen infection, in hybrid necrosis PCD is aberrantly activated leading to the phenotype of a plant responding to pathogen infection but in the absence of a pathogen.

Autonecrosis, a syndrome similar to, but milder than, hybrid necrosis, has been attributed to known *R* genes. For example, in the tomato *Lycopersicon pimpinellifolium*, the gene *Cf-2* confers resistance to the fungus *Cladosporium fulvum* and requires the presence of a second gene, *RCR3*, for its resistance activity. However, if *Cf-2* is combined with the *RCR3* allele from *L. esculentum* (the cultivated tomato) the offspring are resistant to *C. fulvum* but also develop autonecrotic lesions even in the absence of the pathogen. Both genes in this classic DM-like epistatic interaction have been cloned; *Cf-2* encodes a transmembrane protein with extracellular LRRs whilst *RCR3* encodes a cysteine endoprotease. It has been suggested that Cf-2 normally guards Rcr3 and triggers a defence response upon perception of the ligand formed when the Avr2 protein from *C. fulvum* binds to Rcr3, but in the hybrid a subtle structural difference in Rcr3, which differs in seven amino acids between the two species, may result in activation of an

Avr2-independent response upon binding to Cf-2 (Kruger et al., 2002). As stated by Bomblies (2009), this shows the importance of interacting resistance components co-evolving in order to maintain functional interactions and to prevent aberrant defence activation and associated deleterious effects. In addition, when alleles of an *R* gene, *Cf-9*, from another species of tomato were expressed in tobacco, which belongs to the same family as tomato, it resulted in autonecrosis which further suggests that *R* gene-mediated autonecrosis could contribute to barriers between species (Wulff et al., 2004).

The evidence supporting an autoimmune response as the cause of hybrid necrosis continues to grow. In *Arabidopsis*, expression analysis of three genetically independent instances of hybrid necrosis cases were found to involve activation of plant immune genes and necrotic hybrids also exhibited increased pathogen resistance (Bomblies et al., 2007). In one system, the presence of an *R* gene homolog, *Dangerous Mix 1* (*DM1*), from one parent which interacts with an allele at the *DM2* locus of a second strain was both necessary and sufficient for hybrid necrosis. *DM1* shows up to 20% amino acid divergence between three Arabidopsis strains. The characterisation of a further two instances of hybrid necrosis in *Arabidopsis* found the incompatible allele from one strain involved a distinct allele of the same *R* gene cluster as *DM2* (Alcazar et al., 2009). As such, these two systems have allele incompatibilities that map to an overlapping locus but differ in allelic effects and epistatic interactions.

Fine mapping of two hybrid weakness genes in rice has identified *hbd2* as a casein kinase 1 (CKI1) (Yamamoto et al., 2010) and *hbd3* as mapping to an *R* gene cluster. Furthermore, the expression of pathogen response markers was found to be up-regulated in the double homozygote genotype which expresses necrosis despite the absence of a pathogen, suggesting the immune response is active leading to the conclusion that the weakness observed in rice hybrid breakdown can be attributed to an autoimmune response (Yamamoto et al., 2010). In lethal hybrids of a *N. suaveolens* x *N. tabacum* cross, multiple genes related to disease resistance are expressed across the point at which cells were committed to PCD which implies their involvement in the mechanism for hybrid lethality (Masuda, 2007).

Furthermore, Alacazar et al. (2009) demonstrated that the salicylic acid (SA) pathway was important for driving hybrid incompatibility as the extent of the necrosis phenotype correlated with SA flux. Salicylic acid is an endogenous signalling molecule that mediates in plant defence against pathogens by inducing the production of pathogenesis-related proteins (Edreva, 2005, Raskin, 1992). Unlike *R* gene products, the occurrence of these proteins is not pathogen-specific, but determined by the type of reaction of the host plant and their induction is associated with bringing about systemic acquired resistance (SAR). The elevated expression of PR proteins has also been found in necrotic hybrids in tobacco (Mino et al., 2002). As SA-dependent defences are triggered by *R* gene recognition of a pathogen and the HR, it has

fostered the theory that the accumulation of DM incompatibilities leads to SA pathway activation resulting in extensive hybrid necrosis. Alcazar et al. (2009) conclude that multiple loci with small effects, or other complex epistatic networks not yet identified, act as modifiers of the incompatible phenotypes and that different epistatic networks appear capable of promoting hybrid incompatibilities through environmentally conditioned activation of plant immune responses. This implies a potential role for other genes of the plant immune system, and not solely *R* genes, in hybrid necrosis. Additional evidence of the role of plant hormones in hybrid necrosis has been found in *N. suaveolens* x *N. tabacum* hybrids where overproduced ethylene acts as an essential factor mediating PCD and subsequent lethality in this cross (Yamada and Marubashi, 2003) as ethylene-induced genes were found to be expressed at the point where cells are irreversibly committed to PCD (Masuda, 2007).

The suppression of the plant immune system at higher temperatures could vary according to species (Alcazar et al., 2009), which could explain the difference in the temperature thresholds for alleviation of the necrosis phenotype in different genera. Interestingly, there appears in some cases to be a limit on the period of time a hybrid can be exposed to the lower temperature and still have the necrosis phenotype suppressed when placed at a higher temperature. For example, the lethality in hybrid seedlings from a cross between *N. suaveolens* and *N. tabacum*, which progress with features characteristic of programmed cell death (PCD), is only suppressed by temperatures of 30-36^{o}C if the cells are exposed to temperatures of 28^{o}C for less than three hours, longer than this and the cells are irreversibly committed to PCD. PCD during hybrid lethality can be divided into three phases: initiation, commitment and execution suggesting that the factors causing hybrid lethality are expressed three hours after induction of hybrid lethality (Inoue et al., 1996, Masuda, 2007).

There are several features of *R* genes that may predispose them to having a role in incompatibility interactions. First, they experience rapid sequence evolution often under strong selective pressure (Mcdowell and Simon, 2006) frequently in the form of diversifying or balancing selection (Bomblies and Weigel, 2007) and so are highly polymorphic (Ding et al., 2007). Second, they have a high copy number (Bomblies and Weigel, 2007) meaning there is a higher probability of a random mutation occurring within them. Third, the guard-guardee mechanism requires the coevolution of the two partners to ensure appropriate interaction (Bomblies and Weigel, 2007). Fourth, they perform a crucial role in triggering HR and PCD and so their aberrant activity can cause extensive deleterious effects.

This occurrence of several independent cases of tight linkage between hybrid necrosis and the disease resistance response and the implications for a direct role of *R* genes, or similar types of genes, in hybrid necrosis is an interesting point in terms of the evolution of the crossing barrier between Cop and Cerig. Assuming copper tolerance and hybrid necrosis are controlled by

different loci, the tight linkage between them could have caused the hybrid necrosis locus in Cop to have evolved through hitchhiking. However, this may not necessarily be the case if this gene is an *R* gene (that is linked to the copper tolerance gene) as, due to its role conferring resistance to disease, this gene itself is likely to be under direct selection. However, the very tight linkage between necrosis and tolerance (Macnair and Christie, 1983) would make it hard to distinguish whether necrosis evolved via independent selection or hitchhiking.

Thus, in hybrid genomes, aberrant interactions involving *R* genes could potentially cause autoimmune-like incompatibility resulting in PCD and consequently the appearance of the necrotic phenotype (Bomblies and Weigel, 2007). Other cases of hybrid necrosis in which the genetic basis is unknown could involve the immune system, especially if the necrosis is temperature sensitive. The research in this thesis is aiming to elucidate the genetic basis of the crossing barrier between Cop and Cerig and the proposed role of *R* genes in hybrid necrosis adds a fascinating new dimension.

However, it seems improbable that genes of the immune system are the only loci involved in causing necrosis. Indeed, a study by Yamada et al. (1999) found four types of lethality in *Nicotiana*: Type I – hybrid seedlings showed browning of the short apex and root tip; Type II – browning of hypocotyls and roots; Type III - yellowing of true leaves; Type IV – formation of multiple shoots. As these different types of lethality responded differently to the rescue methods of culturing at higher temperature, culturing cotyledonary segments to obtain regenerants, or culturing on a medium containing cytokinin, the authors concluded that hybrid lethality in interspecific crosses of *Nicotiana* differs in the physiological processes that lead to death. In addition, Masuda et al. (2007) identified ninety-nine hybrid-lethality related genes expressed in lethal hybrid cells of a cross between *N. suaveolens* and *N. tabacum*, although the molecular mechanisms and genetic basis of this cross combination remains unknown.

Remarkably, very few genes that have been identified as contributing to hybrid necrosis have been assigned functions other than those involved in the immune response. Jiang, W. et al. (2008) identified a glucose-methanol-choline (GMC) oxidoreductase as the likely product of *Hwh1* and a putative hexose transporter protein as the candidate for *Hwh2* in rice hybrid weakness although could not supply a model for how these could interact to cause hybrid breakdown. However, both these gene products may also be involved in the disease response (Herbers et al., 1996, Lamb and Dixon, 1997). Yamamoto et al (2010) identified *hbd2* as a casein kinase, but this interacts with *hbd3*, which encodes an NBS-LRR gene.
There are examples of other mechanisms by which DM incompatibilities can cause hybrid lethality. For instance, in *Arabidopsis*, Bikard et al (2009) identified a gene duplication event in which the Columbia (Col) and Cape Verde Island (Cvi) accessions have retained alternate functional copies of an essential gene, *HPA*. This gene encodes a histidinol-phosphate amino-

transferase which catalyses an important step in the biosynthesis of histidine. In hybrid progeny homozygous for the non-functional copies, seed development is arrested presumably because the embryo cannot synthesis histidine.

1.5.3 Hybrid lethality genes in Fish

A well documented example of a hybrid incompatibility, and in fact the first hybrid inviability gene to be identified, is the *Xmrk2* locus of *Xiphophorus*. The spotted *X. maculatas* carry a complex known as the *Tumour* (*Tu*) locus that specifies macromelanophores (spots of black-pigmented cells) and is regulated by suppressor locus *R*. However, *X. helleri* lack both the *Tu* locus and dominant alleles at *R*, and as a result some of the hybrids of a cross between these species inherit *Tu* but lack *R* suppressors and so develop lethal melanomas (Coyne and Orr, 2004, Orr et al., 2004, Wu and Ting, 2004). Wittbrodt et al (1989) isolated a tyrosine-kinase encoding gene, *Xmrk2*, from the *Tu* locus and evidence shows that it is responsible for the tumour formation in hybrids. However, based on knowledge from laboratory crosses, Schartl (2008) proposes that hybrid lethality due to malignant melanoma formation is an ineffective barrier to hybridisation. His argument against *Xmrk2* as an effective speciation gene is based on two points; first, that the pigment lesions are never malignant in the F_1 hybrids, which in fact experience hybrid vigour, and second, that in the F_2 only a certain percentage of offspring will develop the lethal melanoma. Yet Schartl (2008) does acknowledge that *Xmrk2* and *R* might have acted as speciation genes in the past and that they may still interact to form a partial postzygotic isolating barrier in the rare instance when pre-zygotic barriers are overcome and an interspecific hybrid formed, two points which re-establish *Xmrk2* and *R* as DM genes involved in speciation.

1.6 Summary and Aims

The research bringing to light information about speciation genes and their vital role in reproductive isolation and speciation is accumulating but despite the advances that have been made over the last twenty years in identifying and characterising these genes, more examples are needed in order to know more fully how widespread these genes are, what functions they perform, the genetic architecture of the incompatibility they contribute to, and how each incompatibility interacts with other isolating barriers to culminate in reproductive isolation and speciation.

This thesis is investigating the partial intrinsic postzygotic isolating barrier between two populations of the yellow monkey flower *Mimulus guttatus* which manifests as hybrid necrosis in the F_1 offspring. Investigating the genetic architecture and characterising the gene(s) responsible for this partial barrier will be a further step towards gaining a deeper knowledge of the genetics of RI and will thus contribute to the understanding of the complexities of speciation based on the following aims:

1) to characterise the hybrid necrosis phenotype further (Chapter 3)

2) to determine whether the copper tolerance gene and the hybrid necrosis gene are synonymous or discrete (Chapter 4)

3) to map the location of the hybrid necrosis locus and identify candidate gene(s) (Chapter 4 & 5)

4) to investigate the small number of polymorphic genes in the Cerig population that interact deleteriously with the single locus in Cop (Chapter 6)

Chapter 2

Methods and Materials

2.1 The study species: *Mimulus guttatus* Fischer ex DC

Mimulus guttatus is a hydrophilic species which in favourable environments can grow as a large herbaceous perennial reaching heights of over a metre whilst in less ideal habitats it grows as smaller facultative annual plants, some as small as five centimetres in arid places (Macnair et al., 1989). The main growing season is April to September and the plants are generally outcrossing with the flowers being bumble bee-pollinated although they are also self-compatible. Figure 2.1 illustrates the main physical features of *M. guttatus*.

Individuals of *M. guttatus* have a principal branching stem that produces multi-flowered inflorescences. The flowers, which vary in size but are of a consistent shape, have fused corollas to which the filaments bearing the anthers are attached, are conspicuous, yellow and dotted with red. The corollas last an average of three to five days after which they abscise with seed capsules developing over eighteen to twenty-two days. In perennial forms, the flowers are slightly protogynous, although phases do overlap, with a sensitive bilobed stigma which closes in less than two seconds when touched and opens within twenty minutes if no pollen has been deposited, an apparent adaptation to insect pollination. The leaves are highly variable in size and shape and are broadly ovate with a dentate edge. Creeping, small-leaved runners are produced in late summer and are easily rooted thereby enabling vegetative spread (www-biol.paisley.ac.uk/bioref/Plantae_Mimulus/Mimulus2).

2.2 Description and provenance of plants used in this research

2.2.1 Copperopolis

At the Copperopolis mine, California (see Figure 2.3), the copper tolerant *M. guttatus* ecotype 'Copperopolis' (see Figure 2.2) grows not only in the streams and damp places but also on the tailings, a dry habitat in which, under normal conditions, it would not grow (Macnair et al., 1989). In these dry sites the plants are small, unbranched annuals producing only a few capsules per plant (Macnair, 1987b, Macnair and Gardner, 1998). However, seed collected from large and small variants produced plants of identical phenotypes when grown under standard conditions

in the laboratory, which implies the smaller size is a response to lack of nutrients and water (Macnair et al., 1989).

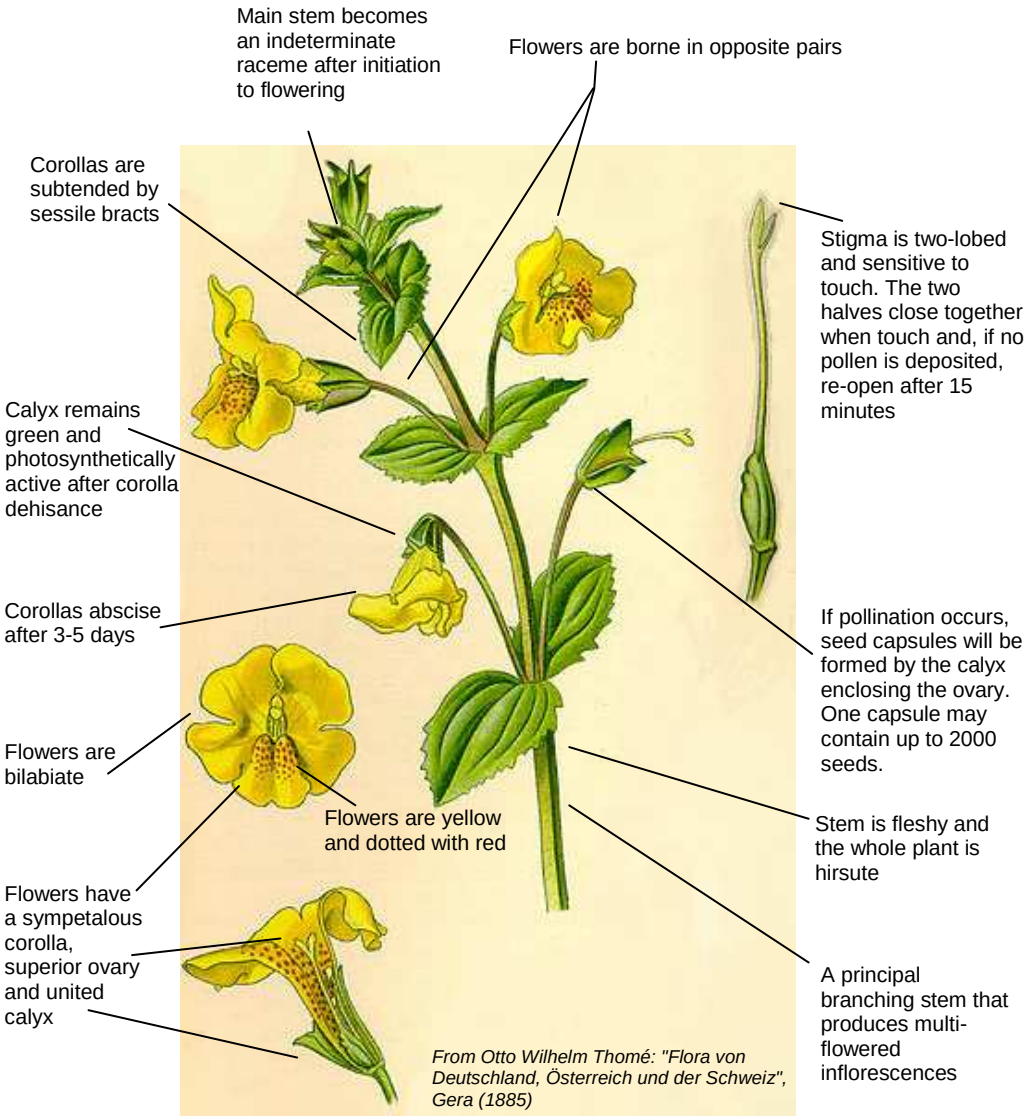


Figure 2.1 The main physical features of *M. guttatus*

The Copperopolis individuals used in this thesis were from seed collected at the Copperopolis mine in 1985 (see Macnair et al., 1993 for details). Seed was collected in bulk from at least fifty plants at a number of different subpopulations. All subpopulations consisted of annual plants: those growing along the streams were large and produced larger numbers of seeds whilst those growing on mine tailings away from the stream were small and produced fewer seeds (Macnair et al., 1989). When grown in the glasshouse for this thesis, the individuals were found to be

significantly smaller than Cerig both in stature and flower size. Furthermore, they tended to behave as annuals making it challenging to maintain the same individuals as perennials for use throughout the duration of the research although some success was achieved by propagating individuals as cuttings.

2.2.2 Cerig-y-Drudion

This is a copper non-tolerant British population from Cerig-y-Drudion (abbreviated to Cerig or Cer) in Conwy, North Wales (see Figure 2.4). *M. guttatus* has been known as a garden escape in the British Isles since 1830 and is now completely feral (Allen and Sheppard, 1971). The original description of *M. guttatus* Fischer ex DC was of plants collected in the Aleutian Islands and, as Cerig certainly has features of the coastal, more northerly populations of this species, it is likely that this population was established by human-assisted dispersal. As such, Cerig resembles a 'typical' *M. guttatus* plant in appearance, possessing a branching stem that gives rise to multiple inflorescences and reaching heights of up to one metre (see Figure 2.2).

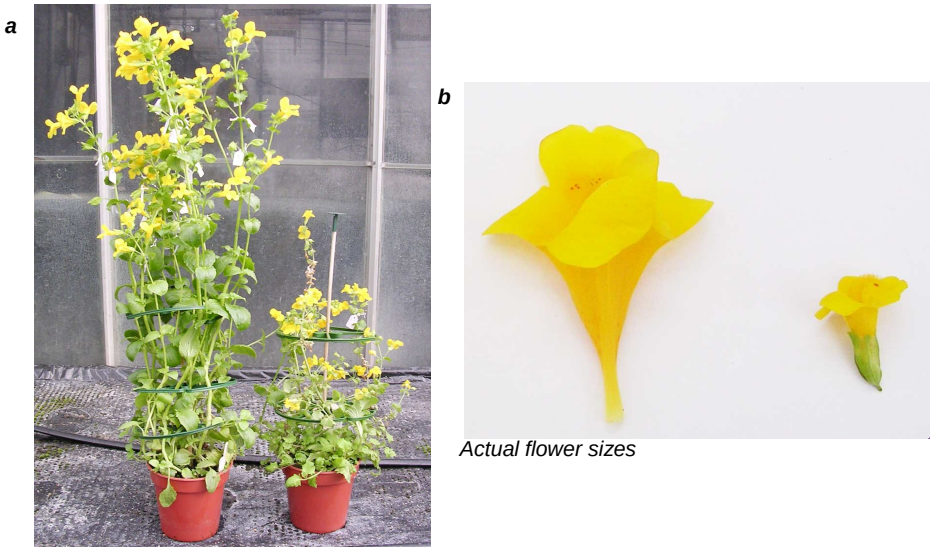


Figure 2.2 The morphological differences between the study populations Cerig (shown on the left) and Copperopolis (shown on the right) showing the differences in (a) height and (b) flower size despite the plants being grown in identical environments.

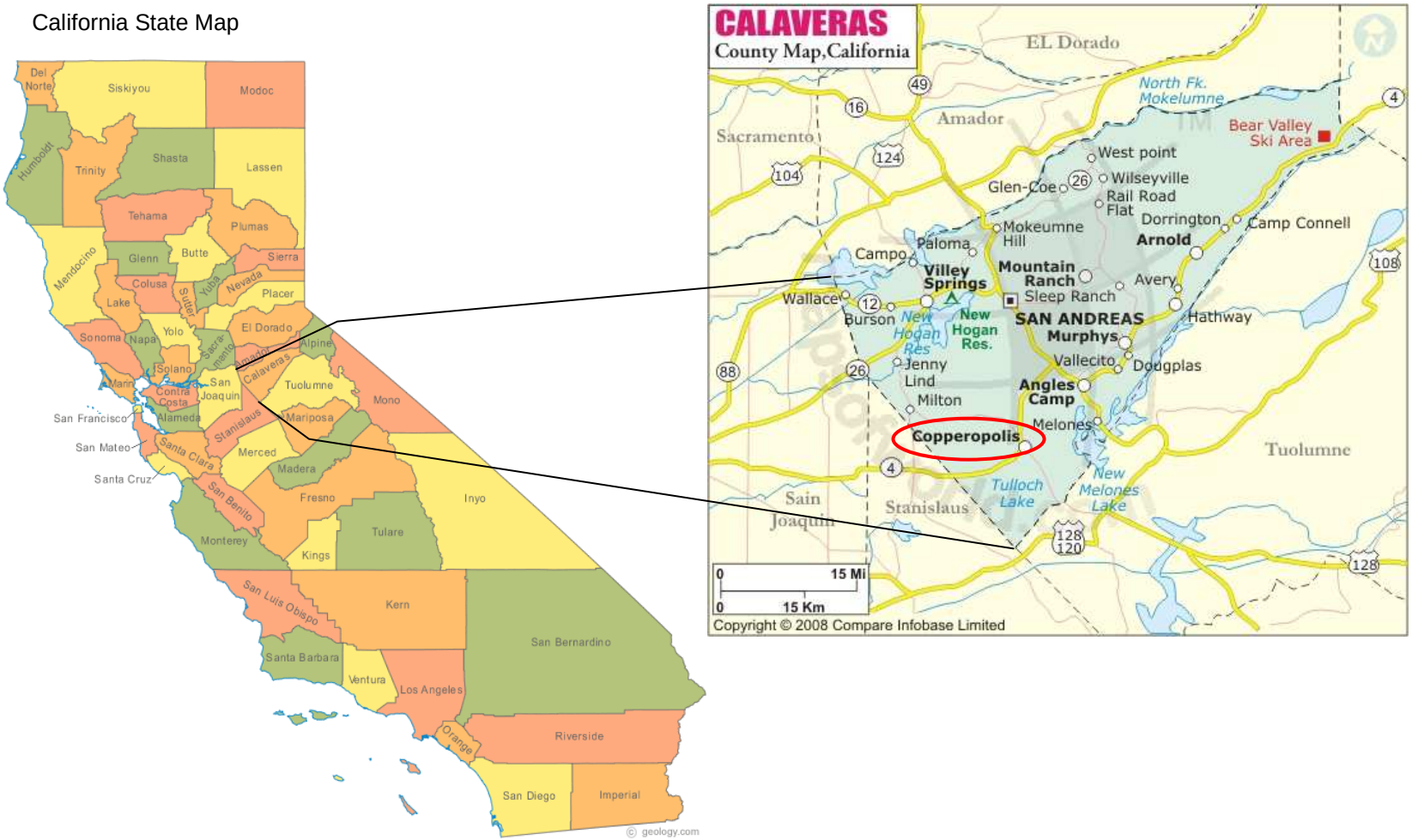


Figure 2.3 Maps from www.geology.com and www.mapsofworld.com showing the location of Copperopolis, Calaveras County, California.

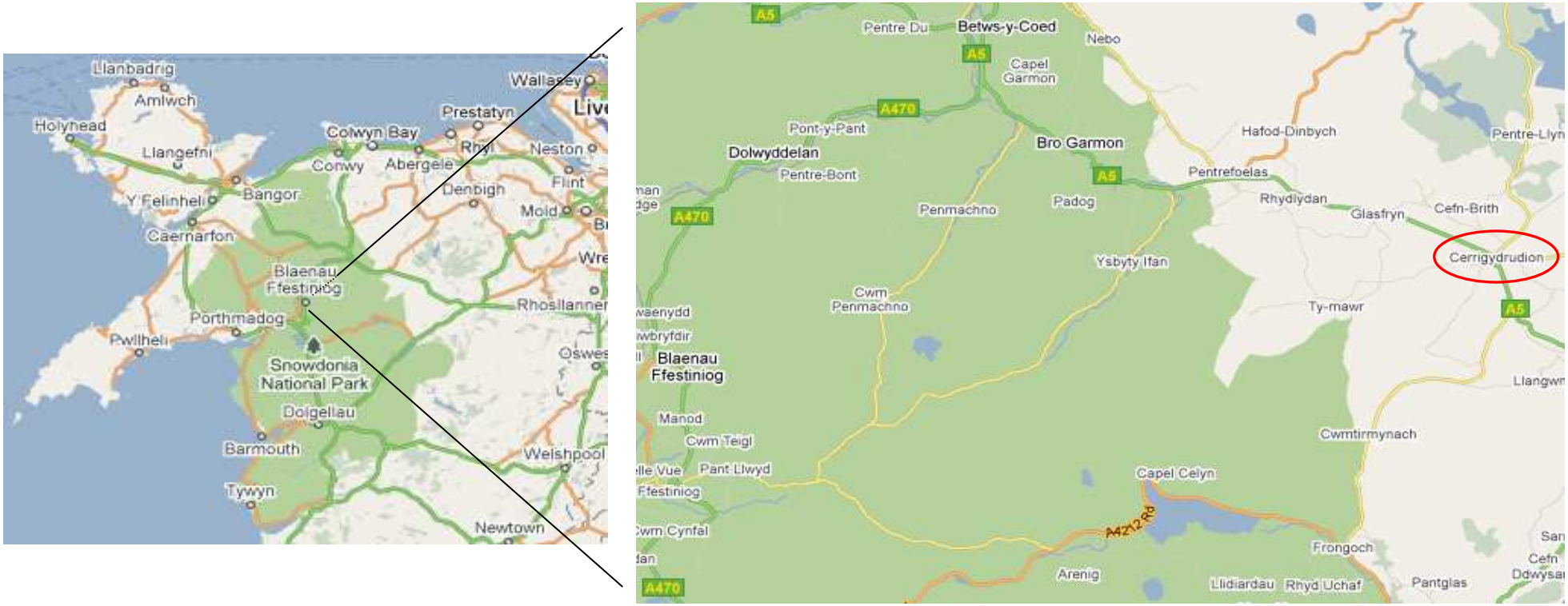


Figure 2.4 Maps from www.maps/google.co.uk showing the location of Cerig-y-Drudion, North Wales, UK.

2.2.3 Plants ‘A’ and ‘D’

These are two plants produced by Professor Mark Macnair by repeatedly backcrossing over five generations an F_1 from a cross between Cop and Cerig to a Stinson Beach (SB) plant, which is a copper non-tolerant population from Marin County, California. As such, both plants share the SB genetic background, but D possesses the copper tolerance locus, and a section of the Cop genome linked to it, giving it the genotype *Tt* and making it copper tolerant whilst A retains the *tt* genotype and so is copper non-tolerant. The genomes of these two plants are therefore likely to be similar except for the region containing the *T* locus.

2.3 Plant growth and care

2.3.1 Raising plants from seed

Seeds were sown in 3½ inch pots containing sieved Levington F_1 seed and modular compost and germinated in a mist unit (see figure 2.5 d). Once large enough to be handled (around 10-12 days after sowing), forceps were used to prick out seedlings into either a medium seed tray (MST) (22cm width x 34cm length x 5cm depth), which held forty-eight plants each, or a small seed tray (SST) (16cm width x 20cm length x 5cm depth), which held thirty plants each, containing a 50:50 mixture of Levington F_1 seed and modular compost and John Innes No. 2. The trays were stood in water until the soil was saturated then placed in the mist unit (see figure 2.5 c) for three to five days prior to being placed in randomised positions on an irrigated bench.

2.3.2 Raising plants from cuttings

Raising plants from cuttings allows the generation of multiple copies of an individual and ensures maintenance of that genotypic line. As well as preventing the risk of losing a genotype through death of a single individual, clones also offer the advantage of enabling the use of multiple plants of an identical genotype during crosses. As such, cuttings were taken of all the parent plants (A, D, Cer10 and Cer35) as well as other important plants (recombinants and segregating F_2s) generated during this project.

Cuttings were taken by using a razor blade to cut branch terminals. The cuttings were rooted in water for around ten days and then potted in to 3½ inch pots containing a 50:50 mix of Levington F_1 seed and modular compost and John Innes No.2 compost. They were then placed in the mist unit for up to a week before being placed on a bench.

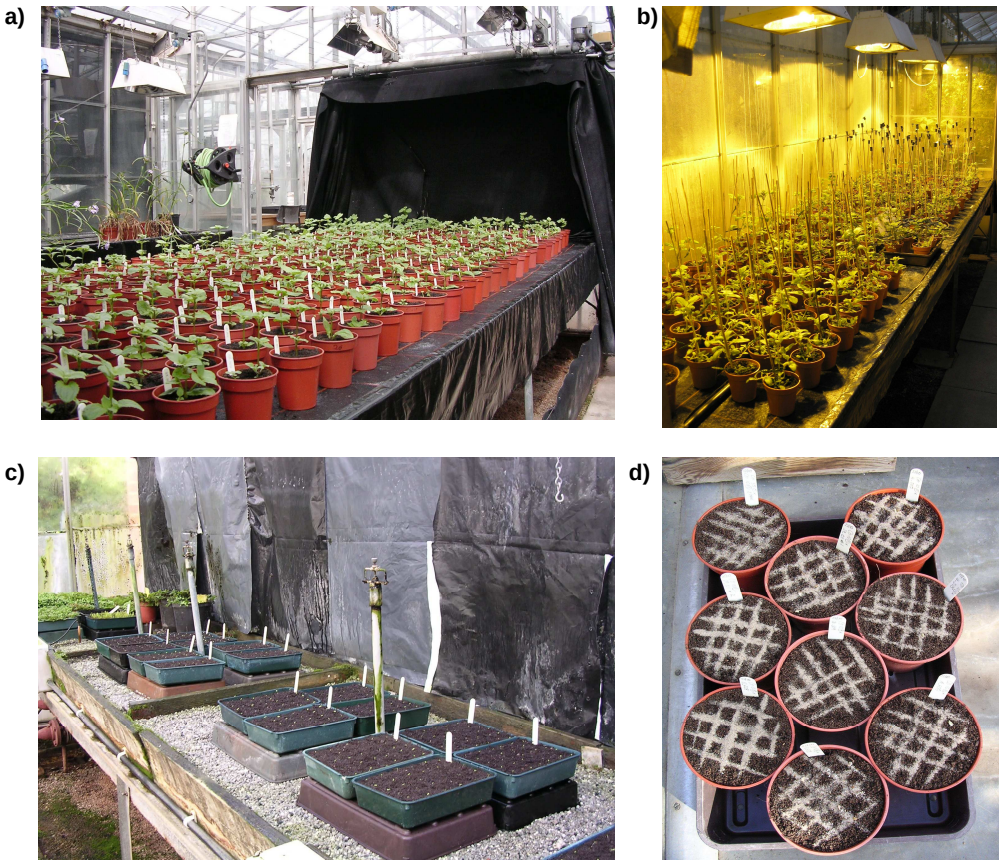


Figure 2.5 Glasshouse plant growth methods (a) plants grown on bench with blackout to prevent flowering; (b) plants grown under sodium lights to promote flowering; (c) newly transferred seedlings in trays in mist unit; (d) pots of newly sown seed.

2.3.3 Soil

The compost used to germinate seeds was Levington F_1 seed and modular compost. A 50:50 mix of Levington and John Innes No. 2 was used to grow seedlings and all other plants including cuttings. This mix provided the best balance of ease of use, nutrients and drainage.

2.3.4 Lighting/blackout

Plants were maintained under short day length conditions (nine hours) by using an automatic blackout screen that opened at 9am and closed at 6pm to promote vegetative growth rather than flowering (see figure 2.5 a). This was particularly important during the process of phenotyping in order to encourage root growth rather than flowering. Days were lengthened to sixteen hours in order to promote flowering. This was achieved either by removal from the blackout or, during darker periods or when natural day length was too short, by artificial lighting using a 400-watt high pressure sodium bulb (see figure 2.5 b).

2.3.5 Temperature

As much as possible, plants were maintained in the same compartment of the glasshouse at a daytime temperature of <20^{o}C and a night time temperature of >12^{o}C. There were, however, fluctuations according to the outside temperature and so glasshouse temperatures could reach as high as 32^{o}C and as low as 8^{o}C depending on the season. Heating and ventilation were used to control the ambient temperature as much as possible.

2.3.6 Watering and feeding

Plants were watered every day by an automatic bench irrigation system. On hot and/or sunny days the plants were given additional water by soaking the bench matting using a hose. Cop individuals were kept in trays and watered by hand less regularly as they were very prone to mildew and root rot when the soil was too wet.

Stock plants were watered with Vitafeed plant food every twelve to sixteen weeks. When required to flower, the relevant plants were watered with Chempak Low Nitrogen Feed No. 8 (NPK Fertilizer) once a week to help promote flowering. The specific content of these feeds is given in table 2.1. Trays of plants that were being grown to score for hybrid necrosis were not given additional feed.

Table 2.1 Specifications of plant feed

Contents	Vitafeed	NPK Fertilizer
Total Nitrogen	19.0%	12.5%
From Nitric nitrogen	5.5%	7%
From Ammoniacal Nitrogen	3.8%	4.5%
From Ureic Nitrogen	9.7%	1.0%
Phosphorus Pentoxide P_2O_5	19%	25%
Potassium Oxide K_2O	19%	25%
Magnesium Oxide (MgO)	-	0.2%
Trace elements:		
Boron	0.013%	0.02%
Copper	0.025%	0.01%
Iron	0.05%	0.2%
Manganese	0.025%	0.02%
Molybdenum	-	0.002%
Zinc	-	0.05%

2.3.7 Seed storage

Seeds were collected into film bags, labelled and stored in plastic containers in the fridge at 10^{o}C.

2.3.8 Pest management

The main pests incurred were fungal infections (primarily mildew), thrips, slugs, caterpillars, snails, whitefly and red spider mite.

Where possible, biological control methods were used to reduce the incidence of pests. Free-standing electric fans were used to increase airflow and reduce humidity in order to reduce the occurrence of mildew, thrips and whitefly. Containers of pheromones were attached to blue and yellow sticky traps and used throughout the canopy of the plants in order to attract and trap thrips and whitefly. The wetness of the soil was closely monitored to avoid over-watering as this increased the risk of fungal infections, especially of the roots. Caterpillars, slugs and snails were picked off by hand.

However, it was also necessary to use chemical treatments in some instances. These treatments included ‘SB Plant Invigorator’, a three-in-one pesticide/ mildewcide /foliar nutrient spray, Bayer Garden ‘Fungus Fighter’, a systemic fungicide spray which could be applied directly and as often as needed to mildew infected plants and as an excellent tool for keeping mildew at bay, Scotts ‘Bug Clear’ insecticide was used on several occasions when there was a severe problem with caterpillars and hand picking them off proved too time-consuming and ineffective, slug pellets were used to reduce slug numbers when they started increasing or, when ineffective, such as during large outbreaks or on snails, the plants were treated with Scotts liquid ‘Slug-It’.

2.4 Glasshouse-based Experimental Techniques

2.4.1 Phenotyping

2.4.1.1 Phenotyping for Copper Tolerance

This method is largely the same as the method described by Allen & Sheppard (1971) and Macnair (1983).

Seedlings were sown and pricked out as described in 2.3.1. Once the plants were tall enough (stems of at least five centimetres), which typically took around eight weeks, the plants were cut off above soil level and all roots, shoots and lower leaves removed. The plants were then cultured in aqueous solution of calcium nitrate (0.5g/l) and copper sulphate (0.5 ppm Cu) for ten days with the solution being changed every second day. The plants were then scored according to whether they had produced roots (tolerant) or not (non-tolerant) (see figure 2.6 (f) and (g)). Plants were cultured in styrofoam coffee cups, which held 170ml of solution, with plastic lids in which six holes were made and labelled 1 - 6 (see figure 2.6 (a, d and e)). For the experiments where a larger number of plants were being phenotyped for copper tolerance, such as in the phenotyping to identify recombinants in which around 4400 plants were tested, plants were cultured in a 30cm x 45cm plastic tray that held seven litres of solution. A 29cm x 44cm polystyrene sheet was cut from a 1 m^2 ceiling tile and twelve rows of six holes made in it (see figure 2.6 (b)). This sheet then floated on the surface of the solution meaning that not only were all cuttings in contact with the solution at all times, even on hot days when the solution in cups can evaporate quickly, but also that changing the solution was quicker and reduced the risk of error arising from replacing the wrong lid on a cup.

2.4.1.2 Phenotyping for Hybrid Necrosis

Seed from crosses were sown and from each one sixty seedlings were pricked out as described in 2.3.1 into two SSTs containing thirty plants each. Using Microsoft Excel to generate random numbers, the trays were randomly arranged on a bench and the plants were left to grow for six to ten weeks depending on the time of year (longer in the early Spring months (ten weeks) and shorter (approximately six weeks) in the late Spring/Summer so as to take in to account the slower plant growth during the darker months. Lights were not used to encourage growth as it could mask the extent of hybrid necrosis either by promoting rapid growth so that plants were scored before those with a late-onset of necrosis began expressing the phenotype or by altering the growing environment, for example by elevating the ambient temperature.

The number of green (healthy) and the number of necrotic (yellow/necrotic/dwarfed/dead) plants in each tray were counted to give a rate of hybrid necrosis. The hybrid necrosis phenotype is described and shown in detail in Chapter 3.

Figure 2.6 Phenotyping for Copper Tolerance. (a) (d) and (e) *M. guttatus* cuttings being phenotyped in polystyrene cups containing a solution of calcium nitrate (0.5g/l) and copper sulphate (0.5 ppm Cu); (b) modification of the cup system to the use of polystyrene ceiling tiles with holes bored in them; (c) a tray of 48 seedlings; (f) a cutting from a copper tolerant plant after culturing for ten days in 0.5 ppm Cu has healthy root production; (g) A cutting from a copper non-tolerant plant after culturing for ten days in 0.5 ppm Cu has no root production and blackened tips of the root primordial.

2.4.2 Crosses/pollination

Mimulus has a closing stigma system. The stigma is two-lipped (see figure 2.7 a) and sensitive so that when it is touched the lips close and the receptive surface is hidden (see figure 2.7 b). If viable pollen has not been deposited on it, the stigma will re-open (see figure 2.7 c) about twenty minutes after being touched. During the period of closure, the stigma cannot be pollinated and as such when it reopens chosen pollen can be deposited by brushing the anthers across it.

Figure 2.7 Photos of a stigma on an *M.guttatus* plant showing the closing stigma system. (a) An unfertilised *M.guttatus* (Cerig) flower with the stigma open. Upon touching with forceps, the stigma closes (b) enabling removal of the anthers by removing the corolla and attached filaments. If the stigma is unfertilised, it re-opens (c) around twenty minutes after being touched.

Pollinations were performed by hand. By using forceps to touch the stigma to cause it to close and then gently pinching at the base of the flower enabling the removal of the corolla and attached filaments, the flower's anthers were removed. The removed flowers were left for twenty minutes for the pollen to dehisce before being brushed over the stigma of the relevant plant. After twenty minutes, if the stigma re-opened the cross was repeated. If the stigma remained closed then the cross was deemed successful and was labelled. After eighteen to twenty days the seed pod was checked to see if it was ripe and the seed collected in a glassine bag, labelled and stored in the fridge.

2.4.3 Tissue Sampling

DNA was extracted from leaf tissue. Leaf tissue samples of about two centimetres^2 were cut from young leaves, folded loosely and placed in a plastic tube on dry ice. Samples that were sent to Duke were placed in tubes in 96-well plates and packed in dry ice for shipping. The scissors used to cut the leaves were cleaned after each sample was taken to ensure no cross-contamination of tissue.

2.5 I.T. Resources

Various I.T. resources were used throughout the project. The primary facility used was the *Mimulus* evolution website (www.mimulusevolution.org) which contains information on the

ongoing *M. guttatus* genome assembly. The GBrowse function on this site enables the search for and inspection of specific scaffolds and gave access to details such as BAC sequences, mRNA sequences, *Arabidopsis thaliana* protein homologs, repeats and markers. The site also contains a Blast function that was used to check primers for sequence alignment similarities other than the desired sequence elsewhere in the genome and a marker search function to identify the location of EST markers. However, this site has been largely superseded by www.phytozome.net website, first available in January 2010. This site provides access to the sequenced and annotated genome and, where possible, each gene has been annotated with protein family database (PFAM), Eukaryotic Orthologous Groups (KOG), Kyoto Encyclopaedia of genes and genomes (KEGG), and Protein Analysis Through Evolutionary Relationships (PANTHER) assignments

2.5.1 Molecular IT Tools

2.5.1.1 Markers

A web-based primer design programme, Primer3 (http://frodo.wi.mit.edu/primer3/), was used to design the primers for new markers. The allele length polymorphism data from marker screening was analysed using the genotyping software GeneMarker®.

2.5.1.2 Whole-gene sequence amplification intended for sequencing

The European Molecular Biology Open Software Suite (EMBOSS) programme RevSeq (http://imed.med.ucm.es/cgi-bin/emboss.pl?_action=input&_app=revseq) was used to convert the mRNA sequence of the candidate gene in to its DNA sequence. EMBOSS align (http://www.ebi.ac.uk/Tools/emboss/align/index.html) was then used to align the DNA sequence of the candidate gene to the BAC sequence spanning that region of scaffold_84 in the *Mimulus* genome assembly. This was verified using NCBI align (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Primer3 was then used to design primers flanking the candidate gene to PCR amplify it for sequencing.

2.5.2 Bioinformatics IT Tools

The National Centre for Biotechnology Information (NCBI) Blast function (http://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to check for protein homologs using the protein blast. The NCBI ORF finder (http://www.ncbi.nlm.nih.gov/projects/gorf/) was used to identify all open reading frames within the genomic region containing the hybrid necrosis gene in Cop.

Along with NCBI Blast, GenScan (www.genes.mit.edu/GENSCAN.html) was also used to identify complete gene structures (i.e. homologs) in the genomic sequence containing the hybrid necrosis gene in Cop. The *Arabidopsis* Information Resource website (http://www.arabidopsis.org/) was used to study the biological functions of and other information on *A. thaliana* protein homologs.

2.6 Molecular techniques

2.6.1 DNA Extraction

Genomic DNA was extracted from leaf tissue using two methods.

2.6.1.1 The CTAB method

This was the principal DNA extraction method used and was based on a CTAB/chloroform protocol (Doyle and Doyle, 1990) modified for use in a 96-well format.

1. Set water bath to 60^{o}C. Remove 96-well plates containing samples from -80^{o}C and place on dry ice. Ensure each tube contains a metal bead.

2. In a fume hood, measure out Cetyltrimethylammonium Bromide (CTAB) extraction buffer and mercaptoethanol using the following formula:

[(number of samples + 10%) x 500µl CTAB] + [(number of samples + 1) x 1µl mercaptoethanol)]

3. Remove the bottom of the 96-well plate and dip the bottom of the tubes in to liquid nitrogen for approximately 15 seconds directly after which clamp the plates into the Genogrinder. Run at 1550-1700 rpm for between 20 seconds - 2 minutes.

4. When tissue is ground to a fine powder, centrifuge down for a few seconds

5. In the fume cupboard, add 501µl of the CTAB + mercaptoethanol mix to each tube. Return plates to Genogrinder for 10 seconds on its lowest setting

6. Incubate the tubes for ~20mins in the 60^{o}C water bath inverting the plates every 5 minutes to mix.

7. Centrifuge the plates for ~10seconds to pellet any remaining solid leaf tissue. Transfer 400µl liquid from each tube to corresponding tubes in a new 96-well plate. Beads can be washed and reused.

8. Add 400µl chloroform to each tube. Return plate to genogrinder for 2-5 minutes on 0.5x rate at 00. Centrifuge for 10mins at 4000rpm. A band of tissue debris will separate the aqueous (upper) and chloroform (lower) layers.

9. Carefully pipette off 280-300µl of the aqueous layer and transfer to corresponding tubes in a new 96-well plate. Avoid drawing up debris of chloroform. If this occurs, re-centrifuge briefly and transfer again.

10. Add an equal volume of cold isopropanol and mix well by inverting. Place in -20^{o}C freezer for 20 minutes to as long as overnight.

11. Centrifuge for 10min at 4000rpm. A greyish gelatinous pellet should form.

12. Pour off the supernatant and use clean paper towels to wick out as much isopropanol as possible. Add 200µl of cold 70% ethanol and flick to mix. Centrifuge at 4000rpm for 10min. Pour off ethanol, wicking as much as possible out with a paper towel. Air dry the pellet overnight in the fume cupboard.

13. Re-suspend DNA pellet in 100µl 1xTE. Flick tube to mix well and replace lids. DNA can be stored in the fridge rather than freezer to avoid constant freeze-thaw.

Recipe:

CTAB Leaf Buffer	500ml	100ml
100mM Tris, pH8.0	1M, 50ml	10mls 1M Tris, pH8.0
1.42 M NaCl	41.5g	8.3g
20mM EDTA	3.72g	0.744g
2% CTAB	10g	2g
2% PVP 40	10g	2g
5mM ascorbic acid	0.44g	0.088g

2.6.1.2 The QIAGEN DNeasy Kit

A Qiagen DNeasy kit was used to extract DNA from a small number of tissue samples. The method given in the handbook that accompanies the kit was used (available online at http://www.qiagen.com/products/genomicdnastabilizationpurification/dneasyplantsystem/dneasyplantminikit.aspx#Tabs=t2).

2.6.2 Primers

2.6.2.1 Primer Design

Primers were designed using the FASTA sequences of mRNAs extracted from the GBrowse function on the www.mimulusevolution.org website. These sequences were copied into Primer3 and the source sequence marked with brackets. The four primer pair sequences returned by the programme were checked for additional sequence alignment similarities in the genome using the Blast function on the www.mimulusevolution.org intranet. The pair with the fewest sequence similarity hits, and therefore the least chance of amplifying an incorrect fragment, were ordered from Invitrogen.

As the cost associated with ordering customized fluorescently labelled primers is significant, a fluorescent labelled primer, M13 (see Schuelke, 2000) was attached to the 5' end of the forward primer in the PCR reactions for identifying new markers in the fine mapping of the 500kb window on Scaffold_84. The PCR is conducted using 3 primers: the unlabelled Forward and Reverse primers specific to the region being amplified and the M13 primer. The M13 sequence is not in the *Mimulus* genome. The first few PCR cycles amplify the region in the genome between the F and R primers. However, as the forward primer was altered to be a chimera between M13 and the *Mimulus* genomic sequence (by adding the M13 tail) the new amplicons will have the M13 sequence incorporated and these are then the target for the M13 labelled primer, and so, after continued PCR amplification, labelled sequences are produced. The primer volumes in the PCR reaction (10µl of M13 labelled F, 1µl of chimera F, 10µl of R) were designed in order to ensure that most of the amplicons are labelled. The capillary sequencer on which the PCR products are sequenced for allele length reads the fluorescent-labelled M13 primer. This M13 sequence 5'-CACGACGTTGTAAAACGAC-3' works for all primer sets.

2.6.2.2 Primer Preparation

Primers were re-suspended in sterile water using the formula:

nmoles (supplied on primer data sheet) x 10 = volume of water to add in µl

Working stocks of primers were prepared as follows:

M13 Primers

The forward and reverse primer pairs were mixed in the following concentrations:

F primer	1 µl
R primer	10 µl
H_2O	89 µl
Total volume:	100 µl

EST Primers

The EST Forward and Reverse primer stocks are kept separately at a concentration of 100 mM. Working stocks contain both primers at a 5mM concentration using the following volumes:

F primer	5 µl
R primer	5 µl
H_2O	90 µl
Total volume:	100 µl

2.6.3 PCR

2.6.3.1 Preparing the Master Mix

These were prepared directly prior to PCR using the following recipes:

Master Mix for M13 Primers (new markers)

	1x
H_2O	3.8 µl
5x Buffer	2.0 µl
$MgCl_2$	0.8 µl
dNTP	0.8 µl
Primer Mix (F 1µM, R 10µM)	0.2 µl
M13 Primer (10 µM)	0.2 µl
Taq Polymerase	0.2 µl
DNA	2.0 µl
Total	10 µl

Master Mix for EST Primers (Duke stock markers)

	1x
H_2O	4.05 µl
5x Buffer	2.0 µl
MgCl	0.8 µl
dNTP	0.8 µl
Primer Mix (from working stock)	0.2 µl
Taq Polymerase	0.15 µl
DNA	2.0 µl
Total	10 µl

2.6.3.2 PCR Programme

The markers were amplified using standard touchdown PCR conditions:

94°C for 5 minutes

- - - - - - - - - - -

cycle 1:

94°C for 30 seconds

58°C for 30 seconds

72°C for 45 seconds

repeat 10 times, reducing annealing temperature from 58°C to 53°C by 1°C at each cycle for the first 5 cycles then maintain at 52°C for the remaining cycles

- - - - - - - - - - -

cycle 2:

94°C for 30sec

52°C for 30sec

72°C for 45sec

repeat 30 times

- - - - - - - - - - -

72°C for 20 min

10°C forever

2.6.4 Gel Electrophoresis

PCR reactions were checked for successful amplification/correct fragment size by running samples on a 1% gel (as shown in figure 2.8 a) made using the following recipe:

1. 1g agarose + 100mls 1xTBE buffer
2. microwave for 2 minutes
3. rinse to cool
4. add 4µl SYBR Safe DNA gel stain (Duke) or EtBr (Exeter)
5. pour in to tray with comb
6. allow gel to set (15 minutes)
7. on parafilm, mix 4-10µl of each PCR reaction with 2µl loading dye
8. load first well of each lane with 5-10µl DNA Ladder (Easy ladder at Duke, Fermentas 1kb ladder at Exeter)
9. load remaining wells with PCR/loading dye mix

The gels were run at 100-120v for between 20-40 minutes according to gel size and size of PCR product and visualised using BioRad UV Gel Imager (see figure 2.8 b). The SYBR Safe DNA gel stain degrades in light so gels were covered with aluminium foil to shield them.

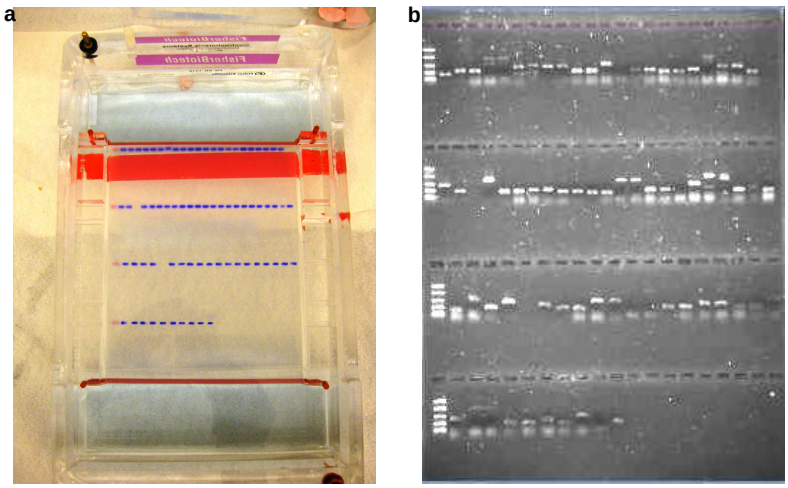


Figure 2.8 (a) Photograph of PCR reactions loaded on a 1% agarose gel prior to electrophoresis; (b) agarose gel visualised using BioRad UV Gel Imager

2.6.5 Genotyping

Markers were based on length polymorphisms in intron regions or microsatellites and genotyping these followed the Willis Lab (Duke) method (Fishman and Willis, 2006, Lowry et al., 2008, Sweigart et al., 2006). Genotyping was performed by subjecting the PCR-amplified DNA

fragments with an incorporated 5' fluorescent-labelled primer to capillary electrophoresis and fragment analysis on an ABI 3730x1 DNA Analyzer (Applied Biosystems). The size of the amplified fragments was scored automatically by the programme GeneMarker (SoftGenetics, 2005, State College, PA) and the score verified by eye.

2.7 Statistical Analysis

The chi-squared and Analysis of Variance statistical tests, along with the Arcsin data transformation method and sequential Bonferroni correction for multiple comparisons, undertaken in this thesis were performed using Excel. Fisher's exact test, used in Chapter 6 as it permits the calculation of precise probabilities in situations where, as a consequence of small cell frequencies, the chi-squared calculations may be inaccurate, was performed using computer software (http://faculty.vassar.edu/lowry/fisher2x3.html).

The standard notation for the levels of significance, with the exception of corrections for multiple comparisons, used throughout this thesis is:

* Significant at $p < 0.05$
** Significant at $p < 0.01$
*** Significant at $p < 0.001$

Chapter 3

Characterising the phenotype of hybrid necrosis in crosses between the Cerig and Copperopolis populations of *Mimulus guttatus*

3.1 Introduction

In *Mimulus guttatus*, the offspring of crosses between a copper-tolerant Californian population 'Copperopolis' (Cop) and certain individuals of a copper non-tolerant British population 'Cerig-y-drudion' (Cerig) exhibit a hybrid lethality phenotype that forms a partial postzygotic intrinsic reproductive isolating barrier (Macnair and Christie, 1983). The phenotype of this crossing barrier is similar to that of hybrids between many other plant species which, due to its characteristic phenotypes of cell death, tissue necrosis, wilting, chlorosis (yellowing), reduced growth rate/dwarfism, and often ultimately death of the plant (Bomblies and Weigel, 2007) is commonly referred to as hybrid necrosis.

As in many other examples of hybrid dysfunctions (sterility or lethality), the hybrid necrosis in crosses between Cop and Cerig is due to the presence of a small number of complementary genes that conform to the Dobzhansky-Muller (DM) model for postzygotic reproductive isolation arising from genic incompatibilities (Macnair and Christie, 1983).

In the simplest two-gene DM model, as described in detail in Chapter 1, an ancestral species with alleles *aa* at one locus and *bb* at another splits into two isolated populations. In each population, the different selection pressures of their distinct environments means that new mutations will be differentially selected for and fixed so that each acquires a new range of unique alleles that the ancestral species did not possess (Coyne and Orr, 2004). As such, a mutation arises in one population causing an *A* allele to appear and become fixed whilst a second mutation, which can occur in either population, causes a *B* allele to appear and also become fixed. These new *A* and *B* alleles are viable in the genetic background of the population in which they arose but may be incompatible with an allele they did not co-exist with in the ancestral genetic background. As a result, if the two diverging populations experience secondary contact and reproduce, incompatible allelic interactions involving the new alleles will result in hybrid dysfunctions in the progeny. Thus, genic incompatibility results from complementary epistasis between loci *A* and *B* in the hybrid genome (Orr, 1995). So, for hybrid

dysfunctions to arise there must be at least two mutation events in either of the lineages and at least one DM gene in each lineage must be involved (Coyne and Orr, 2004, Dobzhansky, 1937, Muller, 1942). Although the DM model of genic incompatibility describes the interaction as occurring between two incompatible genes, it is possible for three or more genes to interact in a similarly complementary manner and result in hybrid dysfunction only when a certain combination of alleles occur (Orr, 1995).

In the initial investigation into the crossing barrier between Cop and Cerig, Macnair and Christie (1983) found that, in Cop, a single locus contributes to the barrier. Furthermore, this locus was found to be the same as, or very tightly linked to, the single major gene for copper tolerance (*T*), but the authors were not able to establish whether the locus contributing to hybrid necrosis was synonymous with or discrete from the copper tolerance locus.

This initial investigation also revealed that the degree of hybrid necrosis (i.e. the proportion of hybrids affected) in this cross varies depending on the individual Cerig parent involved: in some crosses between Cerig and Cop the hybrid necrosis was almost 100% whilst in other crosses it was almost zero, meaning the DM-genes in Cerig responsible for causing hybrid necrosis are polymorphic so that the genotypes *aa*, *Aa* and *AA* all exist within the population with incompatibility occurring between the *A* allele from Cerig and the single locus in Cop (Macnair and Christie, 1983). This effect of polymorphism at DM genes causing variation in the degree of hybrid lethality, ultimately resulting in only partial reproductive isolation, has been found in several other hybrid lethality/necrosis systems in species including *Crepis* (Hollingshead, 1930), *Nicotiana* (Bomblies and Weigel, 2007) and *Gossypium* (Stelly, 1990). In addition, Macnair & Christie (1983) concluded that due to the number of classes of progeny (i.e. the number of classes of necrosis) there must be at least two and probably more genes controlling the Cerig side of the barrier but that the spread of ratios also suggested that there cannot be very many.

As well as the variation in the number of hybrids affected, Macnair & Christie (1983) found that the hybrids varied in the severity of the necrosis phenotype, displaying a range of symptoms, from weak (yellowing) to severe (lethality), an observation that is not uncommon in hybrid dysfunction systems and that has been attributed to a number of causes.

First, the severity of the incompatibility phenotype may depend on the time of onset of the first symptoms, where hybrids affected early in their life tend to suffer the severest phenotype whilst those affected later in life have milder symptoms. This effect was found in the cross between Cerig and Cop (Macnair & Christie, 1983). A variable time of onset of symptoms is not unique to this cross, having also been found in systems of hybrid necrosis in wheat (Hermsen, 1963) and cotton (Song et al., 2009), where the earlier the symptoms of necrosis were manifested, the more severe they ultimately became. As hybrid necrosis is often progressive (Hermsen, 1963) it

is intuitive that the earlier the symptoms appear the more time they have to develop. The cause of the variable time of onset is as yet unknown.

Second, variation in the severity of the hybrid dysfunction phenotype is often attributed to a gene dosage effect in which the number of hybrids affected by the lethal or sterile phenotype increases in proportion to the number of incompatible alleles present. Such dosage effects involve incompatible alleles which are dominant and have been found to cause wide variation in the severity of the phenotype even in seemingly simple, two-gene interactions in *Gossypium* (Song et al., 2009), rice (Ichitani et al., 2007), *Arabidopsis* (Bomblies et al., 2007), and wheat (Chu et al., 2006, Hermsen, 1963) as well as more complex interactions involving more than two loci (Kubo and Yoshimura, 2005). Many of these studies refer to variation in the 'degree' of hybrid necrosis when in fact they are describing variation in the severity of the symptoms. These two terms are not interchangeable as 'degree' refers to the number of hybrids affected whilst 'severity' refers to the nature of the symptoms, i.e. weak (yellowing) or strong (lethal). Furthermore, the term "gene dosage effect" might be more accurately referred to as an "allele dosage effect" as it is the number of dominant incompatible alleles that affect the degree of the incompatibility phenotype and would make this more clearly distinguishable from the involvement of multiple genes in more complex interactions where a gene dosage effect could refer to how many of the multiple genes involved possess an incompatible allele.

Third, variation in the severity of hybrid necrosis can be caused by allelic expressivity, where DM-genes have multiple alleles each of which differ in the degree of sterility or lethality they bring about. For example, in wheat there are three alleles of the hybrid necrosis gene Ne_1 and five alleles of the complementary gene Ne_2. These alleles differ in their degree of expression, ranging from strong to weak, which combined with an allele dosage effect results in an almost continuous distribution of the hybrid necrosis phenotype (Hermsen, 1963).

In addition to the variation in severity that can arise due to the complex nature of the incompatible allelic interactions and genetic polymorphism for the relevant loci, hybrid necrosis has also been found to be affected by environmental factors, particularly by temperature, where the necrosis phenotype is alleviated when hybrids are grown at higher temperatures.

For instance, in intraspecific crosses between *Arabidopsis* strains, Bomblies et al. (2007) found that the symptoms of hybrid necrosis were present 16^{o}C but alleviated when the hybrids were grown at 23^{o}C. Similarly, Alcazar et al. (2009) found that necrosis was alleviated when hybrids were transferred from 14^{o}C to 20^{o}C. As hybrid necrosis was found to involve the activation of genes normally associated with the plants response to pathogen attack and as temperature and humidity influence the activation of plants defences against pathogens (Bomblies and Weigel, 2007, Zhu et al., 2010), the alleviation of the necrosis phenotype has been attributed to

suppression of an autoimmune response at higher temperatures (Alcazar et al., 2009, Bomblies and Weigel, 2007). Likewise, in interspecific crosses in lettuce, Jeuken et al. (2009) found that hybrid necrosis was alleviated at 30^{o}C but reappeared when plants were returned to a 15^{o}C environment and also that the hybrid necrosis locus in one species was the same as a locus for resistance to downy mildew, meaning this is another example of hybrid necrosis caused by the aberrant activation of the immune system whose activity is then suppressed by higher temperatures.

Different species could have different thresholds for immune system activation (Alcazar et al., 2009) meaning the temperature at which necrosis is alleviated may be variable. For example, in tobacco, hybrid necrosis involves immune-related genes but the threshold temperature which must be reached before the phenotype is alleviated is higher than in *Arabidopsis* and lettuce. A similarly high temperature threshold exists in rice, where hybrid weakness between two cultivars is expressed at 28^{o}C but alleviated at 29-30^{o}C (Saito et al., 2007). Interestingly, exposure to a high temperature treatment (34^{o}C) for ten days prior to being grown outdoors (where the temperature was not specified) was sufficient for a small number (eighteen out of eighty-five) of hybrids from this cross to overcome hybrid weakness and produce a next generation. Although these were still weaker than parent plants, this suggests that the repression of the necrosis phenotype by higher temperatures can, in a small number of cases, be sufficient to overcome reproductive isolation.

The ability of higher temperatures to repress the necrosis phenotype can depend on the length of time the hybrid is exposed to temperatures at which the phenotype is expressed. For example, the lethality in hybrids from a cross between *N. suaveolens* and *N. tabacum* is suppressed only if the cells are exposed to temperatures of 28^{o}C for less than three hours, any longer than this and the cells are irreversibly committed to programmed cell death, a hallmark of the plant immune response (Inoue et al., 1996, Masuda, 2007). A similar time limitation is seen in lethal hybrids in some *Gossypium* crosses which grow normally at 39^{o}C but when transferred to a lower temperature become necrotic on a time-scale that is directly correlated to the degree of temperature reduction yet can survive if they are returned to a lethal-repressive environment before the latter stages of lethality become pervasive (Phillips, 1977).

However, this temperature-repression of hybrid necrosis appears to be somewhat cross-specific, as revealed by lethal hybrids from other *Gossypium* crosses that are not temperature-sensitive and crosses among *Nicotiana* species, which vary according to the cross and therefore to the type of lethality (Yamada et al., 1999). In some *Nicotiana* crosses hybrid lethality was seen at 28-30^{o}C but not at 32-38^{o}C in which the hybrids grew to maturity. In another cross lethality was seen at 28-30^{o}C and at 34-38^{o}C but not at 32^{o}C whilst in a different cross there was some delay in the onset of symptoms at 32^{o}C but all hybrids ultimately died.

This inconsistency in temperature-repression of hybrid necrosis could be attributed to different cellular mechanisms by which hybrid necrosis is brought about.

In this chapter, the crossing barrier between plants from the Cerig and Copperopolis populations of *M. guttatus* is described and characterised and plants of contrasting genotype are identified that are used in subsequent studies in this thesis.

3.2 Methods

The techniques and protocols used here are described in greater detail in Chapter 2.

3.2.1 Phenotyping for hybrid necrosis

A hybrid was scored as suffering from hybrid necrosis when it displayed any of the following symptoms: leaf yellowing, browning/necrotic lesions on leaves, dwarfing, or death. As the necrosis symptoms tended to affect the whole plant it eliminated any requirement for a threshold, i.e. a minimum percentage of leaf area/number of leaves on plant affected by necrosis, before a plant was scored as a necrotic hybrid. To determine the degree of hybrid necrosis in a cross, the number of green (healthy) and the number of necrotic (yellow/ necrotic/ dwarfed/ dead) plants were counted.

3.2.2 Establishing polymorphism for hybrid necrosis genes in the Cerig population

Given the polymorphism for the genes in Cerig that contribute to the crossing barrier, it was necessary to establish the degree of hybrid necrosis caused by individual plants. This was achieved by sowing seed from crosses previously performed by Professor Macnair between fourteen fresh plants from the Cerig population, which he collected in 2000, and the plant “D”, which is heterozygous for the copper-tolerance locus. For each cross, thirty seedlings were pricked out and grown for eight weeks then scored for hybrid necrosis.

Once the degree of necrosis caused by each of the fourteen Cerig individuals was determined, further crosses were performed with selected individuals. These crosses were performed in replicate, using multiple clones of the selected Cerig individuals, and reciprocally in order to establish whether there is any parent-of-origin effect on the degree of hybrid necrosis or whether crosses in either direction were analogous. For each cross sown, sixty seedlings were

grown in two trays in a randomly assigned position on a bench and then scored for hybrid necrosis.

3.2.3 Phenotyping for Copper tolerance

The Cerig individuals were all homozygous non-tolerant for copper (*tt*) whilst plant 'D' was heterozygous for copper tolerance (*Tt*). Since all these crosses will segregate approximately 1:1 for tolerance (T): non-tolerance (NT), and necrosis is linked to tolerance, the necrosis phenotype should only affect the copper tolerant plants. In order to test plants to identify any in which the linkage between the tolerance and necrosis phenotypes had broken due to recombination between the loci controlling these phenotypes, it was necessary to establish the copper tolerance phenotype of offspring from crosses between T and NT parent plants. Accordingly, any offspring from the crosses between 'D' and the fourteen Cerigs that displayed hybrid necrosis symptoms were also phenotyped for copper tolerance.

As described in Chapter 2, to establish the copper tolerance phenotype plants were cultured in a 0.5ppm copper solution for ten days and then scored for root growth.

3.2.4 Testing for Temperature/Environmental effects on the degree of hybrid necrosis

A simple experiment was conducted to identify temperature sensitivity in the degree of hybrid necrosis in the cross between Cop and Cerig. Seven crosses between copper tolerant and non-tolerant plants whose repeats had previously been found to have significantly different degrees of hybrid necrosis were sown and thirty seedlings pricked out into four trays. These plants were then grown for twelve weeks under four different conditions: >22^{o}C, >22^{o}C and under lights, <18^{o}C, and normal glasshouse conditions (variable temperature and light intensity) and then scored for hybrid necrosis.

3.3 Results

3.3.1 The hybrid necrosis phenotype

The key characteristics of the hybrid necrosis phenotype include cell death, tissue necrosis, wilting, chlorosis (yellowing), reduced growth rate/dwarfism, and often ultimately death of the

plant (Bomblies and Weigel, 2007). Some of the characteristics affecting the offspring of a cross between Cop and Cerig, including yellowing, necrosis and dwarfism, are shown in figure 3.1. The germination and early growth of offspring from this cross was normal until at least the four leaf stage at which point the first symptoms started to appear in some plants whilst in others the phenotype did not appear until several weeks later. Once the first symptoms appeared, the leaves of affected plants began to yellow and become progressively necrotic. The colour changes from green to yellow to brown (necrotic) in hybrid necrosis marked the gradual degeneration of the chlorophyll apparatus in the leaf cells and the dying of the remaining leaf tissues (Hermsen, 1963).

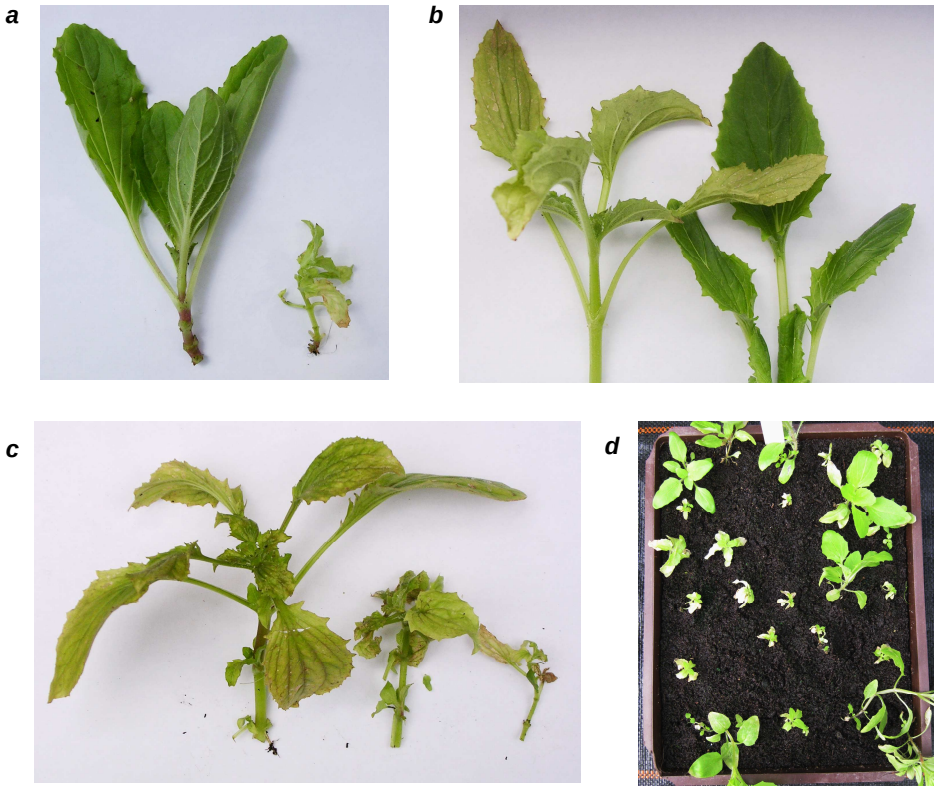


Figure 3.1 The phenotype of hybrid necrosis in progeny from a cross between Cerig (Cer10) and either a copper tolerant ('D') or non-tolerant ('A') plant. (a) A healthy hybrid (left) compared to a necrotic hybrid (right) that exhibits the dwarf characteristic of the phenotype; (b) A necrotic hybrid (left) showing the leaf yellowing/necrosis characteristic of hybrid necrosis and a non-necrotic hybrid (right) which has the green appearance of a healthy plant; (c) Necrotic hybrids can vary in size but all develop yellow necrotic leaves that are often 'crinkled' in appearance. (d) A tray of thirty hybrid offspring where a large proportion (73%) are severely affected by hybrid necrosis exhibiting the yellowing, necrosis and dwarf symptoms.

The severity of the hybrid necrosis phenotype varied widely and appeared to be related to the point at which the first symptoms appeared. Plants affected early in their life usually had severe symptoms and either died or had extensive necrosis and severely stunted growth giving rise to a dwarfed appearance as in figure 3.1 '*a'*. Plants affected later tended to suffer much weaker symptoms, living for many months, mostly exhibiting only the yellowing phenotype and were able to grow as tall as healthy plants, as in figure 3.1 '*b'*, but never reached anthesis. These findings were consistent with those of Macnair & Christie (1983).

In general, the plants with hybrid necrosis were easy to identify as the phenotype was obvious. Even those hybrids suffering only weak symptoms were straightforward to score as the whole plant tended to be affected. There was, however, considerable difficulty in scoring for hybrid necrosis in trays where the plants were, or had been, infected with mildew as these plants also exhibited leaf necrosis which was so similar in appearance to that caused by hybrid necrosis that it was impossible to distinguish between the necrosis caused by genic incompatibility and that caused by mildew. This strongly implies that a similar mechanism may be responsible for hybrid necrosis and for plant resistance to mildew infection, as described in the discussion.

3.3.2 Polymorphism for hybrid necrosis genes in the Cerig population

The degree of hybrid necrosis (i.e. the proportion of hybrids affected) in crosses between Cop and Cerig is known to vary depending on the individual Cerig parent (Macnair and Christie, 1983).

The degree of hybrid necrosis, scored by counting the number of green (healthy) and the number of necrotic (including yellow, dwarf and necrotic) progeny, was established for each of the fourteen Cerig individuals in order to identify the best plant for subsequent crosses.

3.3.2.1 Phenotypic variation in 14 Cerig individuals

The raw data for the degree of hybrid necrosis, given in table 3.1, shows that there is variation in the degree caused by the fourteen Cerig individuals.

In terms of the copper tolerance phenotype of the offspring of these crosses between the Cerig individuals and 'D', no non-tolerant & necrotic hybrids were found, indicating no break in the linkage between tolerance and necrosis. Both tolerant and non-tolerant non-necrotic (green) hybrids were found, as expected considering the polymorphic nature of the Cerig population.

In this initial screen, the individual causing the highest degree of hybrid necrosis was Cerig 10 (Cer10), with a rate of 37.9% whilst those with the lowest hybrid necrosis rates were Cerig 35 (Cer35) at 0% and Cerig 33 (Cer33) at 3.3%. Cerig 34 (Cer34), with a hybrid necrosis rate of 6.6%, had a relatively equal split of copper tolerant and non-tolerant progeny. In order to establish a more accurate average degree of hybrid necrosis for each of these four Cerigs, each was used in crosses with a number of different Cop plants and the offspring scored for the number that suffered from hybrid necrosis.

Table 3.1 Copper tolerance and hybrid necrosis phenotypes in offspring from crosses between different Cerig parents (non-tolerant) and plant D (tolerant).

*G = green (healthy/non-necrotic); L= necrotic (yellow/dwarf/necrotic/lethal); T = copper tolerant; NT = not copper tolerant; * = selected for further crossing experiments with Cop*

Cerig parent plant	G/NT	G/T	L/NT	L/T	Uncertain tolerance phenotype	Total	G	L	NT	T
Cer 10*	16	2	0	4	7	29	18	11	16	6
Cer 14	15	9	0	3	1	28	25	3	15	12
Cer 17	19	4	0	5	0	28	23	5	19	9
Cer 18	17	5	0	1	7	30	22	8	17	6
Cer 31	Not phenotyped for copper tolerance				-	29	29	0	-	-
Cer 32	13	8	0	3	4	28	21	7	13	11
Cer 32.3	Not phenotyped for copper tolerance				-	25	24	1	-	-
Cer 33*	Not phenotyped for copper tolerance				-	30	29	1	-	-
Cer 34*	17	11	0	2	0	30	28	2	17	13
Cer 35*	Not phenotyped for copper tolerance				-	30	30	0	-	-
Cer 36	14	12	0	4	0	30	26	4	14	16
Cer 40	13	11	0	5	1	30	25	5	13	16
Cer 41	15	11	0	2	2	30	26	4	15	13
Cer 42	12	10	0	4	1	27	22	5	12	14

The progeny from thirty-six crosses between seven Cop plants and the four Cerigs identified above were scored for hybrid necrosis. The raw data are given in table 3.2. As all crosses included a Cop parent, which has the copper tolerance genotype *TT*, the progeny were not phenotyped for copper tolerance as they would all be heterozygous with the genotype *Tt* and therefore copper tolerant. Figure 3.2 shows the variation in the degree of necrosis caused by three of the Cerig parent plants.

Table 3.2 Raw data on the degree of hybrid necrosis from crosses between Cerig and Cop plants.

Cross	Healthy	Necrotic	Total	% hybrid necrosis
Cer 10 & Cop				
Cer 10 (11) x Cop (2)	5	55	60	91.67
Cer 10 (2) x Cop (5)	18	42	60	70.00
Cer 10 (3) x Cop (4)	14	46	60	76.67
Cer 10 (3) x Cop (5)	6	54	60	90.00
Cer 10 (6) x Cop (5)	5	55	60	91.67
Cer 10 (4) x Cop (4)	5	55	60	91.67
Cop (1) x Cer 10 (2)	8	52	60	86.67
Cop (4) x Cer 10 (1)	17	43	60	71.67
Cop (4) x Cer 10 (6)	18	42	60	70.00
Cop (5) x Cer 10 (2)	6	54	60	90.00
	102	**498**	**600**	**83.00**
Cer 33 & Cop				
Cer 33 (1) x Cop (5)	**29**	**31**	**60**	**51.67**
Cer 33 (3) x Cop (4)	**41**	**19**	**60**	**31.67**
Cer 33 (3) x Cop (5)	**31**	**29**	**60**	**48.33**
Cer 33 (5) x Cop (7)	**34**	**26**	**60**	**43.33**
Cer 33 (6) x Cop (7)	**44**	**16**	**60**	**26.67**
Cop (1) x Cer 33 (2)	**35**	**25**	**60**	**41.67**
Cop (4) x Cer 33 (3)	**27**	**33**	**60**	**55.00**
Cop (5) x Cer 33 (1)	**31**	**29**	**60**	**48.33**
Cop (5) x Cer 33 (3)	**34**	**26**	**60**	**43.33**
	306	**234**	**540**	**43.33**
Cer 34 & Cop				
Cer 34 (1) x Cop (4)	49	11	60	18.33
Cer 34 (3) x Cop (5)	42	18	60	30.00
Cer 34 (5) x Cop (2)	59	1	60	1.67
Cer 34 (5) x Cop (7)	48	12	60	20.00
Cop (2) x Cer 34 (5)	55	5	60	8.33
Cop (4) x Cer 34 (1)	51	9	60	15.00
Cop (4) x Cer 34 (3)	18	12	30	40.00
Cop (5) x Cer 34 (4)	42	18	60	30.00
	364	**86**	**450**	**20.41**
Cer 35 & Cop				
Cer 35 (1) x Cop (4)	55	5	60	8.33
Cer 35 (3) x Cop (4)	60	0	60	0.00
Cer 35 (4) x Cop (4)	58	2	60	3.33
Cer 35 (2) x Cop (4)	29	1	30	3.33
Cop (4) x Cer 35 (2)	50	10	60	16.67
Cop (4) x Cer 35 (3)	60	0	60	0.00
Cop (5) x Cer 35 (4)	52	8	60	13.33
	364	**26**	**390**	**6.43**

Figure 3.2 Variation in the degree of hybrid necrosis in three trays containing offspring from crosses between a Cop individual and three different Cerig individuals. The variation in the degree of hybrid necrosis is evident, ranging from 83% (Cer10) in the tray on the left, 50% in the middle tray (Cer33) and to 0% in the tray on the right (Cer35).

The hybrid necrosis data from these crosses between Cop plants and the four Cerigs were tested for homogeneity between reciprocal crosses and between groups of replicate crosses for each Cerig parent using a contingency chi-squared test.

3.3.2.2 Testing for homogeneity between reciprocal crosses

The results of a chi-squared test for homogeneity between reciprocal crosses are given in table 3.3. There was no significant difference in the rate of hybrid necrosis depending on the direction of the cross in crosses involving Cer10 and Cer33. There was a significant difference depending on the direction of the cross in crosses involving Cer34 and Cer35 but the χ^2 is small in comparison to values seen when comparing hybrid necrosis between the Cerig parents. The crosses in which a significant difference was found according to the direction of the cross were both of overall lower hybrid necrosis and had the fewest replicates. As such, it is possible that these factors caused errors in scoring for the hybrid necrosis phenotype. Given the low χ^2 value, it was concluded that any parent-of-origin effect on the degree of hybrid necrosis would seem to be very minor compared to the effect of the Cerig individual (see section 3.3.2.4).

Table 3.3 Chi-squared (χ^2) values for the difference in the degree of hybrid necrosis in reciprocal crosses between Cerigs and Cop

(= significant at p<0.05, df = 1)*

Cross	% Hybrid Necrosis	χ^2
Cer10 x Cop	85.28	3.309
Cop x Cer10	79.58	
Cer33 x Cop	40.33	2.474
Cop x Cer33	47.08	
Cer 34 x Cop	17.50	5.002*
Cop x Cer34	26.19	
Cer35 x Cop	3.81	5.969*
Cop x Cer35	10	

3.3.2.3 Testing for homogeneity amongst crosses within each group

As there was considered to be no parent-of-origin effect, the hybrid necrosis data from reciprocal crosses was pooled to give one value for each cross between Cop and either Cer10, Cer33, Cer34, or Cer35. These are given in Table 3.4.

As multiple plants were used to perform the crosses, a contingency chi-squared test was carried out to test for homogeneity amongst the crosses within each group according to which Cerig parent was used. Table 3.5 shows the chi-squared values.

Table 3.4 The degree of hybrid necrosis in crosses between Cop and either Cer10, Cer33, Cer34 or Cer35

Cerig Parent	No. of healthy hybrids	No. of necrotic hybrids	Total no. of hybrids scored	Range in degree of hybrid necrosis (min % – max %)	Average % hybrid necrosis
Cer10	102	498	600	70 - 91.67	83.00
Cer33	306	234	540	26.67– 55.00	43.33
Cer34	364	86	450	1.67 – 40.00	20.41
Cer35	364	26	390	0 - 16.67	6.43

Table 3.5 Chi-squared (χ^2) values for testing for homogeneity amongst crosses within each Cerig x Cop group

*(significant at: * p < 0.05; ** at p < 0.01;*** at p < 0.001)*

Cerig Parent	Range in degree of hybrid necrosis	χ^2
Cer 10	70 - 91.67% (21% spread)	35.861 (9df)***
Cer 33	26.67 – 55% (28% spread)	16.425 (8df)*
Cer 34	1.67 – 40% (38% spread)	34.702 (7df)***
Cer 35	0 – 16.67% (17% spread)	24.375 (6df)***

Within each group of crosses there is significant difference in the degree of hybrid necrosis. This is unexpected considering the fact that the multiple plants within each of the groups are actually cuttings originally obtained from a single plant and therefore genetically identical. However, these are relatively small significance values when compared to the χ^2 values seen when the different Cerig parents are compared (see below) and can be explained by various other factors causing necrosis, such as environmental conditions or error in scoring for necrosis. Alternatively, there could be a small effect of the Cop parent causing variation in hybrid necrosis. In order to limit any potential effect of different Cop parents on hybrid necrosis, a limited number of Cop individuals were used in all the crosses performed in this project. It was, however, necessary to use more than one Cop individual in order to maximise the number of flowers available to use in crosses and also, because Cop plants tend to be more annual, it can be difficult to maintain them for long periods of time or get individuals to flower repeatedly. Ultimately, the number of Cop individuals was constrained by propagating those used by cloning in order to maintain the use of the same plants. However, the key point is that the χ^2 values in table 3.5 are relatively small.

3.3.2.4 Differential necrosis depending on the Cerig parent

Comparing the hybrid necrosis between all four of the Cerigs using chi-squared gives a very highly significant value of χ^2 = 707.97 (p <0.001, 3df) confirming that there is polymorphism in the Cerig parent for the incompatible loci that contribute to hybrid necrosis.

As shown in Table 3.2, the degree of hybrid necrosis ranges from 0% to ~92%. Examining the average rates of each Cerig parent it appears that Cer10 gives a much higher rate than the other Cerigs. Testing whether Cer10 causes a significantly higher rate of hybrid necrosis than the other Cerigs can be achieved, albeit rather crudely, by using four separate chi-squared tests

in which homogeneity between all crosses is tested and then one Cerig at a time omitted from the calculations. The results of this are shown in Table 3.6 and show how much more drastically the significance of the difference in the degree of hybrid necrosis between the Cerigs drops when Cer10 is removed from the calculations than when any of the other Cerigs are omitted.

Table 3.6. Chi-squared (χ^2) values for testing for homogeneity in the degree of hybrid necrosis amongst crosses with data from each Cerig being omitted sequentially. All χ^2 values are significant *** at p<0.001.

Data from…	χ^2
All four Cerigs	707.97 (3df)***
Cer10 omitted	**174.69 (2df)*** **
Cer33 omitted	708.97 (2df)***
Cer34 omitted	563.81 (2df)***
Cer35 omitted	441.74 (2df)***

This confirms that Cer10 causes the highest degree of hybrid necrosis in crosses with Cop and that this rate is significantly higher than the degree of hybrid necrosis in crosses between Cop and the other three Cerigs. Conversely, crosses between Cer35 and Cop results in a lower rate of hybrid necrosis than any of the other Cerigs. This confirms the findings from table 3.1 meaning, of the individuals in the Cerig collection available for crossing experiments, Cer10 and Cer35 cause the extremes of the degree of hybrid necrosis in that Cer10 gives a high average hybrid necrosis rate of 83% whilst Cer35 gives a low average of 6.67%.

3.3.3 Hybrid Necrosis and Temperature

The simple experiment to identify an effect of temperature on hybrid necrosis in the cross between Cop and Cerig by growing replicate trays of crosses under four different conditions was inconclusive. There was no significant difference in the rate of necrosis between plants grown at >22^{o}C and those grown at >22^{o}C under lights, with necrosis ranging between 0-20% in the >22^{o}C group and 0-33% in the >22^{o}C + lights group. Although three out of the seven crosses had a significantly different degree of necrosis between trays grown at >22^{o}C and those grown at <18^{o}C, there was no significant difference (χ^2 = 3.242, p<0.05, 1df) in the average degree of necrosis between the high temperature group and the low temperature group which implies there is no temperature effect. However, there was a significant difference (χ^2 = 45.635, p<0.001, 1df) when the necrosis data from the group grown at <18^{o}C was compared with the

data from group grown in the >22^{o}C + lights conditions, perhaps because these were actually consistently in excess of 26^{o}C.

However, these results are questionable as unforeseen technical problems in the glasshouses meant the temperature was not as tightly controlled as it needs to be to accurately assess the effect of temperature on necrosis. There was a wide fluctuation in temperature in the <18^{o}C treatment, with the plants experiencing temperatures as high as 24^{o}C which consequently means there was no consistent difference in the high and low temperature treatments which would mask any temperature effect. In addition, there were only a small number of crosses sown and for each cross only one tray of thirty offspring was placed in each of the different treatments. Furthermore, as necrosis in *Arabidopsis* is alleviated at 20^{o}C (Alcazar et al., 2009) the temperature for the low temperature treatment might have been too high, even if it had remained consistently at <18^{o}C, to see an accurate effect.

As such, the results from this small trial to test for an effect of temperature on the degree of hybrid necrosis do indicate that the phenotype is temperature sensitive but are inconclusive and can neither prove nor disprove a temperature effect. However, given the evidence from examples in the literature of the prevalence of a temperature effect and the strong possibility for the involvement of the immune system in the Cerig x Cop necrosis, it remains likely that there is an effect of temperature on the degree of hybrid necrosis.

3.5 Discussion

The incompatibility phenotype in the hybrid offspring of a cross between a copper tolerant Copperopolis plant and copper non-tolerant individuals from the Cerig population has the features characteristics of hybrid necrosis including tissue necrosis, chlorosis (yellowing), reduced growth rate and death and varies in both severity and degree.

The severity of the necrosis phenotype appears to depend on the time of onset of the first symptoms – the earlier the first symptoms are manifested, the more severe the phenotype is and results in greatly stunted growth, with plants reaching little more than a few centimetres in height, severe necrosis and ultimately death. Later onset of the first symptoms is associated with a weaker phenotype, usually involving only yellowing with no effect on growth and with affected plants living for several months although never reaching anthesis. A similar effect of time-of-onset has been found in other cases of necrosis such as in *Gossypium* (Song et al., 2009). However, the severity of the phenotype resulting from the various crosses was not recorded as the focus of this research project is the degree of hybrid necrosis.

The degree of hybrid necrosis ranged from 0 - 91.67% depending on the Cerig parent, confirming Macnair & Christie's (1983) finding that the Cerig population is polymorphic for a small number of loci involved in an incompatibility interaction with the single tolerance-linked locus in Cop. As there are a small number of polymorphic genes involved in the Cerig side of the RI barrier it is possible that there is an allele dosage effect meaning the strength of the barrier (i.e. the number of necrotic hybrids) depends on the arrangement of alleles at these Cerig loci. However, although unproven by this characterisation work, it is also likely that the degree of necrosis is influenced by temperature, which could explain the variation in the degree of hybrid necrosis between replicate crosses shown in table 3.5, and as such the ambient temperature of the growing environment is an important factor when growing plants to score for the proportion of necrosis.

A very strong similarity was found between the appearance of hybrids with necrosis caused by genic incompatibilities and those with necrosis resulting from mildew infection and most likely therefore activation of the immune response. This is in keeping with several examples in the literature in which these similarities have led to the suggestion that epistatic interactions between incompatible alleles can trigger aberrant activation of the plant immune system (Bomblies and Weigel, 2007). Studies of hybrid necrosis have found the involvement of genes associated with pathogen response, including several examples of disease resistance genes (*R* genes) in tobacco (Masuda, 2007), *Arabidopsis* (Alcazar et al., 2009, Bomblies et al., 2007) and rice (Yamamoto et al., 2010), as well as genes encoding proteins that interact with *R* gene products (Kruger et al., 2002, Jeuken et al., 2009, Wulff et al., 2004), genes that encode other factors in the disease response (Herbers et al., 1996, Lamb and Dixon, 1997), and plant hormones associated with the disease response (Alcazar et al., 2009) and programmed cell death (Yamada and Marubashi, 2003). As such, an autoimmune response brought about by incompatibilities involving immune-system genes may be a common molecular mechanism underlying hybrid necrosis across diverse species (Bomblies et al., 2007). The similarity between the appearance of plants with hybrid necrosis and plants with a mildew infection along with the growing literature on this similarity strongly suggests the possibility that the hybrid necrosis in crosses between Cop and Cerig could involve genes that function in the plant's immune system.

Whatever the mechanism causing hybrid necrosis, the outcome of the polymorphism in Cerig is that reproductive isolation (RI) between these two populations is variable and not complete and the strength of the crossing barrier (i.e. the number of necrotic hybrids) depends on the alleles at the Cerig loci.

3.6 Conclusion

Having characterised the hybrid necrosis phenotype in the offspring of crosses between Cop and Cerig and identified plants of contrasting genotype for use in the subsequent studies, this research project can now focus on two objectives: identifying the single locus in Cop and beginning a preliminary investigation into the loci in Cerig that are collectively responsible for causing hybrid necrosis in crosses between these populations.

Chapter 4

Identifying the genomic region containing the hybrid necrosis (*NEC*) locus in Copperopolis

4.1 Introduction

The offspring from crosses between two populations of *Mimulus guttatus,* a copper-tolerant Californian population 'Copperopolis' (Cop) and certain individuals of a copper non-tolerant British population 'Cerig-y-drudion' (Cerig), exhibit a hybrid necrosis phenotype. This intrinsic postzygotic reproductive isolating barrier has been attributed to the presence in both populations of a small number of incompatible genes that complement each other in a Dobzhansky-Muller type interaction (Macnair and Christie, 1983).

Research by Macnair and Christie (1983) on the incompatibility between the Cop and Cerig populations established that a single gene in Cop interacts with a small number of polymorphic genes in Cerig to cause hybrid necrosis in their progeny. Furthermore, this gene in Cop, referred to as the hybrid necrosis (*NEC*) locus, was demonstrated to be very tightly linked to a locus that confers copper tolerance (Macnair and Christie, 1983). Copper tolerance has evolved in some populations of *M. guttatus*, including Cop, that grow on abandoned copper mines in California (Allen and Sheppard, 1971) and is due to the presence of a single major copper tolerance locus (*T*) (Macnair, 1981, 1983).

It has not yet been established whether *NEC* is a gene very tightly linked to the *T* locus or is the copper tolerance gene itself. Although no recombinants between the *T* and *NEC* loci were found by Macnair & Christie (1983) during their research, the authors were only able to screen around one-hundred potentially recombinant plants and so may not have found recombinants as a consequence of a limited sample size. Indeed, the major impediment to establishing whether or not the *T* and *NEC* loci in Cop are synonymous or discrete has been the substantial requirement such a project would have in terms of space and manual labour. Large-scale phenotyping for copper tolerance, although time-consuming, is not impossible as each plant can be phenotyped directly by culturing a cutting in an aqueous copper solution. However, to establish the phenotype of a plant in terms of its propensity to cause hybrid necrosis necessitates performing crosses in order to generate its hybrid progeny and then growing up a sufficient number (at least sixty) of these hybrids to score the number that suffer necrosis. This is enormously time-consuming and requires an extensive work area and so it is not feasible to

test many thousands of plants in an attempt to identify recombinants between the *T* and *NEC* loci.

However, because of the very tight linkage between the *T* and *NEC* loci, the probability of finding a recombinant between these loci can be enhanced by identifying and using closely linked markers for the *T* locus. With this strategy, plants in which recombination has occurred between the *T* locus and linked markers can be identified and then only these need to be tested for their hybrid necrosis phenotype. If the relationship between copper tolerance and hybrid necrosis has dissociated, revealed by a copper tolerant parent causing low hybrid necrosis or a copper non-tolerant parent causing high hybrid necrosis, then the recombination event must have disrupted the linkage between the *T* and *NEC* loci meaning that these loci are discrete. If there are markers both up- and downstream of the locus, knowledge of marker genotypes and copper tolerance phenotype can then enable the identification of a recombination window in which the *NEC* gene is located.

Assuming the *T* and *NEC* loci are discrete, the rate at which the linkage between the copper tolerance and hybrid necrosis phenotype is disrupted by recombination will be a direct result of the distance between them. The rule is that the greater the distance between two loci, the greater the possibility for recombination to occur between them. As such, if the *T* and *NEC* loci are located very close together, recombination between the *T* locus and a marker is less likely to occur in-between the *T* and *NEC* loci and so the linkage between the copper tolerance and hybrid necrosis will largely remain intact giving rise to mostly tolerant-necrotic and non-tolerant-non-necrotic progeny. If the distance between the *T* and *NEC* loci is large then the chance of recombination between the *T* locus and a marker occurring in-between the *T* and *NEC* loci is more likely and so there will be a larger number of progeny in which the copper tolerance and hybrid necrosis phenotypes dissociate giving rise to the recombinant phenotypes of tolerant-non-necrotic or non-tolerant-necrotic. Whether there are few or many, the identification of plants in which recombination has occurred between the *T* and *NEC* loci will provide the first evidence that these are discrete loci. However, given the very tight linkage between them (Macnair and Christie, 1983), it is likely that if the *T* and *NEC* loci are discrete that they exist in close proximity and so there are unlikely to be many recombinants between them. This renders the identification and use of the *T* locus linked markers to screen a large numbers of plants of great value.

The aim of this section of research is to identify *T* locus linked markers and use these to elucidate more about the nature of the *T* and *NEC* loci as to whether they are synonymous or discrete and to ultimately determine a limited region in which the *NEC* locus is located. This requires the identification of plants where recombination has occurred between the *T* locus and linked markers, as the genotypes and associated phenotypes at marker alleles will allow the

identification of a recombination window. By crossing these recombinant plants to a Cerig plant that causes high hybrid necrosis and calculating the hybrid necrosis rate in the offspring, it will be possible to test to see whether recombination between the *T* locus and its markers causes a break in the linkage between the copper tolerance and hybrid necrosis phenotypes.

4.2 Methods and Materials

The techniques and protocols used here are described in greater detail in Chapter 2.

4.2.1 Identifying Recombinants

This was carried out in collaboration with Kevin Wright at Duke University, North Carolina who undertook the molecular work (marker screening and genotyping).

The *M. guttatus* plants screened for recombination were the progeny grown from seed generated by crossing two introgression lines: A, which was copper non-tolerant, and D, which was copper tolerant (see 2.2.2.3). By using these plants, differences in the degree of hybrid necrosis in progeny from the crosses detailed in 4.2.3 and 4.3.1 can be attributed to the section of the genome that contains the copper tolerance locus and consequently also the *NEC* locus.

4.2.1.1 Phenotyping for Copper tolerance

The F_1 seedlings from a D x A cross were grown for six to eight weeks after which leaf tissue samples were taken and sent to Duke for marker genotyping and the plants were phenotyped for copper tolerance by culturing in an aqueous solution of calcium nitrate (0.5g/l) and copper sulphate (0.5ppm Cu) for ten days and scoring for root production.

4.2.1.2 Genotyping

Identifying Markers

As the plants A and D are very closely related, having been produced by repeated backcrossing into a non-tolerant genetic background (Stinson Beach), their genomes are likely to be similar except for the region around the tolerance locus. This was used as a basis for identifying markers for the copper tolerance locus. The markers were based on expressed sequence tags

(ESTs), or sequenced mRNAs, generated by Clemson University, and at a later date subsequent markers were generated using the JGI *M. guttatus* genome sequence.
A total of 541 co-dominant markers developed to target intron length polymorphism in *M. guttatus* were tested in the A and D lines. Markers that were polymorphic were then tested in 72 F_1 D x A progeny to identify those strongly associated with the copper tolerance genotype. This was performed by Kevin Wright and a summary of the results is given in table 4.1.

Table 4.1 A summary of the results of the marker screening in plants A and D and D x A F_1 progeny conducted by Kevin Wright at Duke University to identify markers linked to copper tolerance

Total number of marker screened in A and D	541
Of these, the number found to be heterozygous in the D line	52
Of these, the number linked to tolerance (screened in D x A F_1)	4
Name and scaffold location of the four markers linked to tolerance	
MgSTS217	Sc_68: 778382 – 780439
MgSTS242	Sc_177: 505167-50619
MgSTS536	Sc_68: 250794 – 252493
MgSTS745	Sc_124: 623334 – 625424

Identifying Recombinants

The process of genotyping the D x A plant material sent from Exeter involved DNA extraction using the CTAB method followed by a PCR reaction with primers specific for the markers, submission of the PCR product to the Duke Genomics Facility for fragment analysis to determine the fragment size and analysis of the data received from the genomics facility using GeneMarker software.

The marker MgSTS242 (e242), identified as being linked to the tolerance locus (see 4.3.1), was used to identify individuals in a D x A F_1 population in which recombination had occurred between it and the *T* locus. In order to identify plants in which this recombination event had occurred, 3486 progeny from a D x A cross were successfully scored for their copper tolerance phenotype and genotyped for the e242 marker. More plants (a total of 4381) than the 3486 included in the analysis were grown and screened but were ultimately excluded from the analysis primarily because they could not be assigned a definite genotype due either to noise/stutter or failed fragment amplification with PCR or because of ambiguity in assigning their copper tolerance phenotype.

The additional markers described in 4.3.2 were not identified until after completion of the crosses between the recombinants and the selected Cerigs. As such, these markers were not used to identify recombinants for use in crosses but were critical in identifying the recombination window containing the *NEC* locus.

4.2.2 Selection of Cerig individuals to use in testcrosses with recombinants

Individuals within the Cerig population show polymorphism for hybrid necrosis (Macnair & Christie, 1983) and the work on characterising the hybrid necrosis phenotype (Chapter 3) included identification of the degree of necrosis caused by multiple Cerig individuals. For use in crosses with recombinants, the Cerig individual that contributed to the highest degree of hybrid necrosis was selected (Cer10). As a control, the individual that gave the lowest degree of necrosis was also selected for crossing to the recombinants (Cer35).

4.2.3 Crosses between Cerigs and recombinants

A total of sixty-eight plants were identified by the work in 4.2.1 as being recombinants between the *T* locus and e242. Of these, twenty-seven plants were used in crosses with Cer10. Crosses were performed between Cer10 and both copper tolerant recombinants (TR) and copper non-tolerant recombinants (NTR) by hand pollination and the progeny from all these crosses were sown and scored for hybrid necrosis.

To reduce the effects of potential data inaccuracies arising from death in the hybrid progeny caused by factors other than hybrid necrosis, crosses were performed in replicate with each recombinant plant. Crosses were also performed reciprocally so that any parent-of-origin effect would be revealed. As a control, crosses between Cer35 and five TRs and three NTRs were performed and the progeny scored for hybrid necrosis with the expectation that necrosis would be low given that Cer35 causes low hybrid necrosis in crosses with Cop.

Hybrid necrosis data were collected from a total of 129 crosses, including replicates, between Cer10 and the recombinants. Of these, sixty-two were crosses between Cer10 and the TRs and sixty-seven were crosses between Cer10 and the NTRs. With the exception of five crosses in which, due to low seed number/germination, only thirty progeny were grown and one cross in which only forty-five progeny were grown, sixty progeny from each cross were grown and scored for the number of necrotic hybrids.

4.2.4 Phenotyping for hybrid necrosis

Hybrid necrosis was determined by counting the number of green (healthy) and the number of necrotic (including lethal, dwarf and yellow) progeny for each cross sown to give a proportion of hybrids suffering necrosis.

4.2.5 Calculating Recombination Frequencies

The recombination frequency (RF) between the tolerance locus (*T*) and the first marker identified (e242) was calculated by dividing the number of recombinants by the total number of plants scored for that marker. The RF calculation for markers and loci other than e242 used the formula *X/N x (RF at e242)*, where *X* is the number of recombinants between the *T* locus and the marker, and *N* is the number of recombinants between e242 and *T* tested. This formula only works, of course, if the order of loci is e242 – marker – *T* locus.

4.3 Results

4.3.1 Testing the linkage between the T and NEC loci

As the *T* and *NEC* loci are tightly linked, identifying recombinants between them by scoring many plants for both phenotypes is not feasible. However, identifying plants in which markers for the *T*-locus have recombined with the *T*-locus and phenotyping only these plants for their necrosis phenotype in order to find recombinants between *T* and *NEC* is more feasible. Plants which possessed a copper tolerant phenotype but had a marker genotype normally associated with non-tolerance were labelled tolerant recombinants (TRs). Conversely, those that were copper non-tolerant but had a marker genotype normally associated with tolerance were labelled non-tolerant recombinants (NTRs).

At the time of conducting the crosses between Cer10 and recombinant plants, the tolerance locus marker that had been identified as lying closest to the *T* locus was MgSTS242 (e242) and so all the recombinant plants were derived by the genotyping of this marker to identify recombination between it and the *T* locus.

Marker e242 is heterozygous and so has two alleles that vary in length - one allele is 168 base pairs (bp) and the other is 178bp. Progeny from a D x A cross usually had either the genotype *168/178* and were copper tolerant or the genotype *178/178* and were copper non-tolerant. This means that the 168bp allele is linked to the *T* allele of the copper tolerance locus and the 178bp

allele is linked to the *t* allele. However, recombination between the *T* locus and e242 resulted in tolerant plants with the genotype *178/178* and non-tolerant plants with the genotype *168/178*.

A total of sixty-eight recombinant plants were identified using e242 and although it would have been preferable to use all of these in crosses to Cer10, only twenty-seven could be used due either to plant mortality or to delayed identification meaning the plants had either been discarded or were identified too late to be included in crosses as the Cerig parents had ceased to flower. Of these twenty-seven plants, eleven were TRs and sixteen were NTRs. The mean degree of hybrid necrosis associated with each recombinant parent plant is shown in figure 4.1 and an additional summary of the data is given in table 4.2. The raw data from each cross is supplied in Appendix 4.1.

Table 4.2 The number of replicates scored for hybrid necrosis for each recombinant (between marker e242 and the T locus) parent plant in crosses with Cer10 and the range in the degree of hybrid necrosis within these replicates.

Plant	Number of Replicates	Range in degree of hybrid necrosis (%) between replicates
TR		
25_A09	4	51.67 – 80.00
25_E11	**6**	**6.67 – 26.67**
26_C07	8	35.00 – 51.67
26_H02	8	26.67 – 63.33
28_G06	4	38.33 – 55.00
32_G02	8	51.67 – 70.00
34_A03	8	35.00 – 66.67
35_F09	4	45.00 – 73.33
43_G01	4	43.33 – 71.67
43_H01	4	46.67 – 65.00
53_B01	4	48.33 – 56.67
NTR		
10_F08	4	1.67 – 6.67
14_D05	4	0.00 – 11.67
19_C03	4	0.00 – 15.00
19_F06	5	1.67 – 15.00
19_H12	3	5.00 – 10.00
24_D09	5	1.67 – 13.33
25_E01	**6**	**31.67 – 65.00**
29_G11	4	0.00 – 13.33
32_H09	6	0.00 – 11.67
35_A12	6	0.00 – 13.33
35_B06	2	0.00 – 30.00
41_C05	5	0.00 – 11.67
43_B01	3	1.67 – 10.00
43_C01	4	1.67 – 6.67
46_B12	2	0.00 – 6.67
48_B01	4	1.67 – 13.33

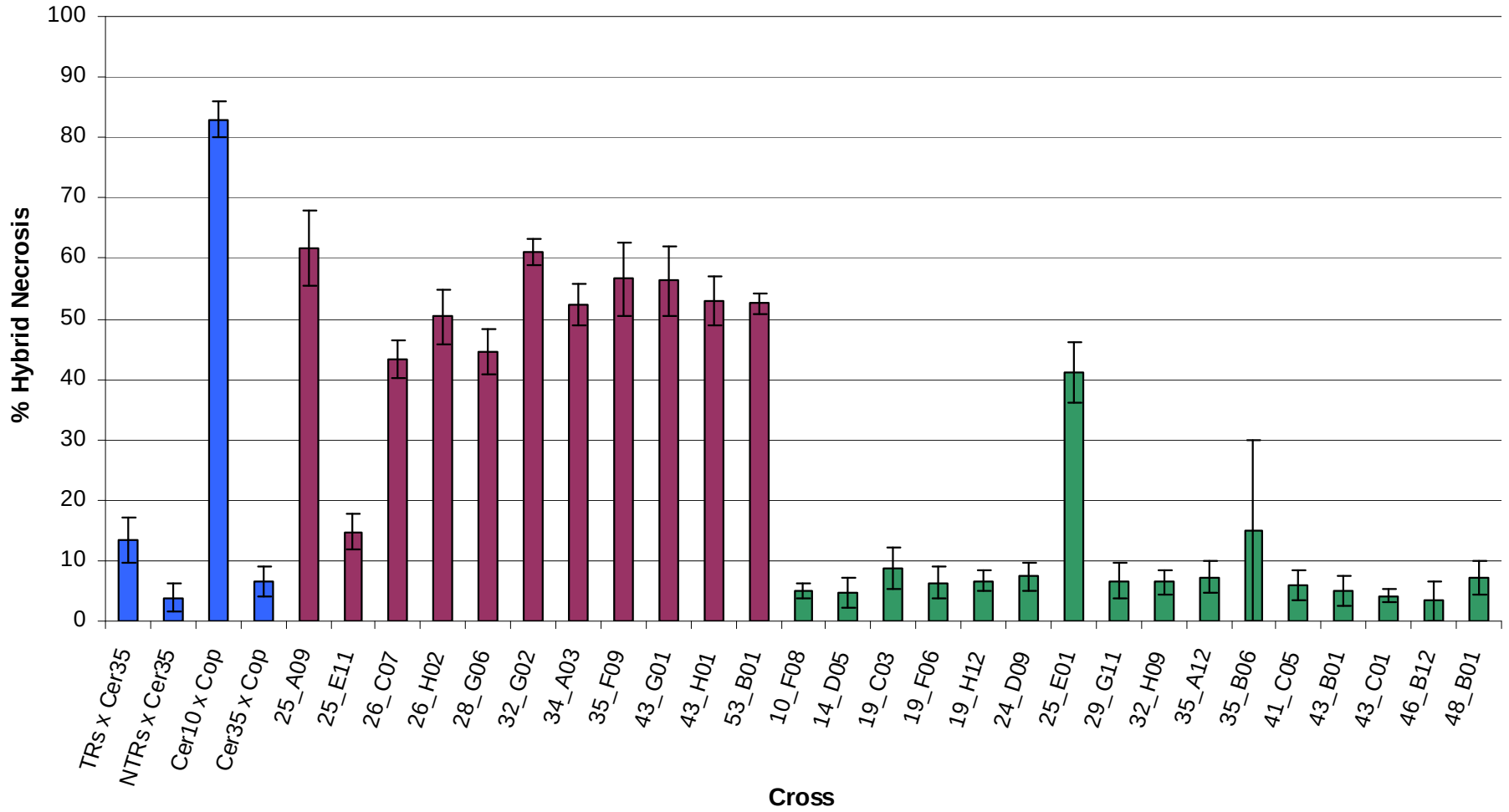


Figure 4.1 The mean degree of hybrid necrosis in control crosses (blue) between grouped TRs or NTRs x Cer35, Cer10 x Cop and Cer35 x Cop and in testcrosses between Cer10 and each individual tolerant recombinant (TR) (purple) or non-tolerant recombinant (NTR) (green).

4.3.1.1 Testing for homogeneity amongst repeat crosses

As replicate crosses were performed with each recombinant plant, a contingency chi-squared test was conducted to test homogeneity within each set of replicates. The results are shown in tables 4.3 and 4.4.

Table 4.3 Chi-squared (χ^2) values for comparing the degree of hybrid necrosis in replicate crosses for each tolerant recombinant

(*significant at * p<0.05, ** p<0.01, *** p<0.001*)

TR	df	χ^2 value
25_A09	3	11.986**
25_E11	5	12.467*
26_C07	7	12.76
26_H02	7	27.835***
28_G06	3	4.098
32_G02	7	6.86
34_A03	7	13.890
35_F09	3	6.774
43_G01	3	10.074*
43_H01	3	4.867
53_B01	3	0.869

Table 4.4 Chi-squared (χ^2) values for comparing the hybrid necrosis data from replicate crosses with each non-tolerant recombinant

(*significant at * p<0.05, ** p<0.01, *** p<0.001*)

NTR	df	χ^2 value
10_F08	3	2.105
14_D05	3	10.195*
19_C03	3	10.401*
19_F06	4	14.272**
19_H12	2	1.607
24_D09	4	8.633
25_E01	5	18.404**
29_G11	3	9.545*
32_H09	5	11.379*
35_A12	5	18.240**
35_B06	1	21.167***
41_C05	4	11.771*
43_B01	2	4.912
43_C01	3	2.087
46_B12	1	4.138*
48_B01	3	6.882

Although the degree of hybrid necrosis in replicate crosses with four TRs and ten NTRs are significantly different, the χ^2 values are not particularly high in comparison to those obtained

when comparing other crosses, for example the χ^2 value obtained when comparing the hybrid necrosis resulting from crosses between Cer10 and the TR or NTR (see table 4.7). These differences might result from errors in hybrid necrosis scoring or from environmental factors such as an inconsistent ambient greenhouse temperature or mildew infection affecting the accuracy in the scoring for hybrid necrosis and the difference in hybrid necrosis could become not significant if more repeats were conducted to reduce the effects of these factors on the data. As such, the hybrid necrosis data were combined to produce an average degree of hybrid necrosis for each recombinant (shown in figure 4.1).

4.3.1.2 Testing for a parent-of-origin effect

Although it was concluded in Chapter 3 that there was no parent-of-origin effect on hybrid necrosis, reciprocal crosses were conducted in order to ensure that this remained the case between Cer10 and the recombinants. As shown in table 4.5, the degree of hybrid necrosis in the grouped TR & Cer 10 and the grouped NTR & Cer10 crosses is significantly different depending on the direction of the cross, which implies a parent-of-origin effect. However, the differences are not uniform in direction: there is higher hybrid necrosis when Cer10 is the maternal plant in crosses with TRs but the converse is true in crosses with NTRs when higher hybrid necrosis is found when Cer10 is the paternal parent. Furthermore, a significant difference in the degree of hybrid necrosis does not occur unanimously amongst the recombinants as reciprocal crosses in sixteen recombinant individuals have no difference in their degree of hybrid necrosis. In addition, although these differences in the degree of hybrid necrosis in reciprocal crosses should not be neglected, as with the Cerig x Cop versus Cop x Cerig crosses (see Chapter 3), these values are comparatively small and it is more likely that the variation arose from a source other than a parent-of-origin effect, such as errors in phenotyping or the effect of the growing environment. Taking into account the data from the crosses between Cerig and Cop and these other factors, it was concluded that there was no parent-of-origin effect on hybrid necrosis and so the data from reciprocal crosses with each TR or NTR could be pooled.

4.3.1.3 Testing the effect of a TR or NTR parent on the rate of hybrid necrosis

The research into the crossing barrier between Cop and Cerig has consistently found that copper tolerance is associated with a high rate of hybrid necrosis whilst non-tolerance is associated with a low rate of hybrid necrosis (Macnair & Christie, 1983; Chapter 3 of this thesis). Whether this holds true in the progeny of crosses between the recombinants and Cer10 was tested by using chi-squared to compare the differences in hybrid necrosis data depending on whether Cer10 was crossed with a TR or an NTR. This gave a very highly significant result of

Table 4.5 Chi-squared (χ^2) values for comparing the mean hybrid necrosis in reciprocal crosses between Cer10 and grouped or individual TRs and NTRs (*significant at 1df * p<0.05, ** p<0.01, *** p<0.001*)

Cross	Mean hybrid necrosis (%)	χ^2
TRs x Cer10 **Cer10 x TRs**	**51.02** **46.84**	**6.357***
25_A09 x Cer10 Cer10 x 25_A09	68.33 55.00	4.510*
25_E11 x Cer10 Cer10 x 25_E11	12.00 16.11	1.130
26_C06 x Cer10 Cer10 x 26_C06	46.25 40.42	1.660
26_H02 x Cer10 Cer10 x 26_H02	56.67 44.17	7.500**
28_G06 x Cer10 Cer10 x 28_G06	44.44 45.00	0.010
32_G02 x Cer10 Cer10 x 32_G02	61.25 60.83	0.010
34_A03 x Cer10 Cer10 x 34_A03	55.71 47.50	3.030
35_F09 x Cer10 Cer10 x 35_F09	54.44 54.17	0.000
43_G01 x Cer10 Cer10 x 43_G01	55.00 57.50	0.150
43_H01 x Cer10 Cer10 x 43_H01	57.50 48.33	2.020
53_B01 x Cer10 Cer10 x 53_B01	52.50 52.50	0.000
NTRs x Cer10 **Cer10 & NTRs**	**7.51** **11.29**	**15.950*****
10_F08 x Cer10 Cer10 x 10_F08	4.17 5.83	0.350
14_D05 x Cer10 Cer10 x 14_D05	7.50 1.67	4.670*
19_C03 x Cer10 Cer10 x 19_C03	7.50 8.89	0.130
19_F06 x Cer10 Cer10 x 19_F06	2.50 8.89	4.950*
19_H12 x Cer10 Cer10 x 19_H12	5.00 7.50	0.400
24_D09 x Cer10 Cer10 x 24_D09	3.33 10.00	4.710*
29_G11 x Cer10 Cer10 x 29_G11	1.67 11.11	8.510**
25_E01 x Cer10 Cer10 x 25_E01	52.50 42.50	3.020
32_H09 x Cer10 Cer10 x 32_H09	10.00 4.58	3.920
35_A12 x Cer10 Cer10 x 35_A12	9.17 6.25	1.020
35_B06 x Cer10 Cer10 x 35_B06	0.00 30.00	21.180***
41_C05 x Cer10 Cer10 x 41_C05	0.00 9.84	11.210***
43_B01 x Cer10 Cer10 x 43_B01	2.50 10.00	4.740*
43_C01 x Cer10 Cer10 x 43_C01	2.50 5.83	1.670
46_B12 x Cer10 Cer10 x 46_B12	3.33 -	-
48_B01 x Cer10 Cer10 x 48_B01	2.50 11.67	6.410**

χ^2 = 1428.96 (significant at p<0.001, 1df) meaning that, as expected, there is a considerable difference in hybrid necrosis depending on whether the recombinant parent was copper tolerant or non-tolerant.

The summarised data in table 4.6 clearly shows that this large χ^2 value is the consequence of hybrid necrosis being much higher in crosses between Cer10 and the TRs than with the NTRs. In the Cer10 and NTR crosses, out of a total of 3885 hybrid plants just 9.55% showed the hybrid necrosis phenotype whereas in the Cer10 and TR crosses out of the 3630 plants scored, 48.98% showed necrosis. This provides further verification of the very strong association between the hybrid necrosis phenotype and copper tolerance.

Table 4.6 Pooled data for hybrid necrosis in TR and NTR crosses with Cer10

Cross	Number of healthy hybrids	Number of necrotic hybrids	Total number of hybrids scored	Mean hybrid necrosis (%)
Cer 10 & TR	1852	1778	3630	48.98
Cer 10 & NTR	3514	371	3885	9.55

4.3.1.4 Identifying recombinants between the T and NEC loci

The pooled hybrid necrosis data for the separate TR and NTR groups was tested for homogeneity using chi-squared. Table 4.7 shows the χ^2 values amongst the crosses in the TR & Cer 10 group and the NTR & Cer 10 group. These very highly significant χ^2 values mean that the rate of hybrid necrosis is not homogenous within the TR and NTR groups. Although some significant difference was seen in the crosses between Cer10 and Cop in Chapter 3, the χ^2 value was much smaller at just χ^2 = 35.86 which, when combined with the fact that all the Cer10 individuals used in crosses with recombinants originated from one single plant, means the large variation in hybrid necrosis can only be due to the different recombinant parent plants.

Table 4.7 Chi-squared (χ^2) values for testing for homogeneity of hybrid necrosis rates within the TR or NTR groups in crosses with Cer10 *(significant at: * p<0.05; ** p< 0.01; *** p< 0.001)*

Cer10 & …	df	χ^2 value
TR	*61*	*326.89****
NTR	*66*	*635.26****

To discern the source of this large difference in hybrid necrosis within each group, the average hybrid necrosis rate for each individual recombinant was studied. From the graph in figure 4.1, two recombinants, TR 25_E11 and NTR 25_E01, stand out as having notably lower and higher hybrid necrosis rates respectively than the other recombinants in their groups.

As TR 25_E11 is copper tolerant it is expected to cause a high rate of hybrid necrosis yet only 14.72% of the offspring were affected. The average rate of necrosis in the other ten TRs is between 43.33% - 61.67%. If the chi-squared test is repeated with the necrosis data omitted from each TR in turn (i.e. a reduction in the heterogeneity with one degree of freedom) the χ^2 value is affected as shown in table 4.8. When TR 25_E11 is omitted the χ^2 value drops by 181.09. Omission from the chi-squared test of any of the other TRs has a much less drastic effect on the value with the reduction in χ^2 varying between 2.06 when TR 53_B01 is omitted and 37.9 when TR 32_G02 is omitted.

This means that in TR 25_E11 the tolerance phenotype is no longer associated with the hybrid necrosis phenotype meaning that the linkage between the *T* and *NEC* loci has been disrupted by the recombination event between the *T* locus and marker e242 indicating the discrete nature of these loci.

Table 4.8 Chi-squared (χ^2) values from tests where the necrosis data from each TR in turn is omitted (NB. *All χ^2 values are significant at p<0.001 where df † =10, †† = 9)*

Data from…	χ^2	Reduction in χ^2
All TRs	326.89†	-
25_A09 omitted	299.32††	27.57
25_E11 omitted	**145.80**††	**181.09**
26_C07 omitted	307.17††	19.72
26_H02 omitted	303.72††	23.17
28_G06 omitted	320.78††	6.11
32_G02 omitted	288.99††	37.9
34_A03 omitted	311.97††	14.92
35_F09 omitted	317.76††	9.13
43_G01 omitted	311.70††	15.19
43_H01 omitted	320.53††	6.36
53_B01 omitted	324.83††	2.06

The difference in the NTRs can be tested in the same way as with the TRs. As NTR 25_E01 is non-tolerant, it is expected to cause a low rate of necrosis. The average rate of necrosis in the other fifteen NTRs is between 0 – 15 % but 41.11% of the hybrid progeny scored from crosses between NTR 25_E01 and Cer10 suffered symptoms of necrosis. If the chi-squared test is repeated with the necrosis data omitted from each NTR in turn, the χ^2 value, as shown in table 4.9, decreases the most when NTR 25_E01 is omitted.

As with TR 25_E11, these data collectively mean that in NTR 25_E01 the non-tolerant phenotype has become associated with the hybrid necrosis phenotype indicating that the linkage between the *T* and *NEC* loci has been disrupted by the recombination event between the *T* locus and e242. As with the TR analysis, this reveals that the *T* and *NEC* loci are discrete.

Table 4.9 Chi-squared (χ^2) values from tests where the necrosis data from each NTR in turn is omitted. *All χ^2 values are significant at p<0.001 where df † =15, †† = 14*

Data from …	χ^2	Reduction in χ^2
All NTRs	635.26†	-
10_F08 omitted	610.89††	24.37
14_D05 omitted	604.36††	30.9
19_C03 omitted	620.93††	14.33
19_F06 omitted	606.33††	28.93
19_H12 omitted	624.11††	11.15
24_D09 omitted	615.94††	19.32
25_E01 omitted	183.81††	**451.45**
29_G11 omitted	612.98††	22.28
32_H09 omitted	604.59††	30.67
35_A12 omitted	605.19††	30.07
35_B06 omitted	609.68††	25.58
41_C05 omitted	608.74††	26.52
43_B01 omitted	615.33††	19.93
43_C01 omitted	605.69††	29.57
46_B12 omitted	616.76††	18.5
48_B01 omitted	616.42††	18.84

Although the degree of hybrid necrosis associated with 25_E11 is closer to the degree normally found with the NTRs than the other TRs, and similarly the necrosis rate of 25_E01 is closer to that of the TRs than the other NTRs, there is, interestingly, heterogeneity in the degree of hybrid necrosis between both these recombinants and the other recombinants of similar phenotype. The degree of hybrid necrosis caused by 25_E01 is still significantly lower than that of eight of the TRs, as shown in table 4.10, whilst the necrosis rate of 25_E11 remains significantly higher than that of fourteen of the NTRs, as shown in table 4.11.

Table 4.10 Chi-squared (χ^2) values from tests comparing the degree of hybrid necrosis of each TR, and their group average, with 25_E01.

Degree of hybrid necrosis with	Compared to 25_E01
25_A09	24.340***
26_C06	0.420
26_H02	7.160**
28_G06	0.710
32_G02	32.770***
34_A03	8.390**
35_F09	9.270**
43_G01	13.240***
43_H01	8.080**
53_B01	7.530**
Average TRs	8.860**

Table 4.11 Chi-squared (χ^2) values from tests comparing the degree of hybrid necrosis of each NTR, and their group average, with 25_E11.

Degree of hybrid necrosis with	Compared to 25_E11
10_F08	12.790***
14_D05	14.180***
19_C03	4.640*
19_F06	10.480***
19_H12	6.530**
24_D09	7.690**
29_G11	9.590**
32_H09	11.650***
35_A12	8.970**
35_B06	0.040
41_C05	10.170**
43_B01	10.180**
43_C01	15.670***
46_B12	10.420**
48_B01	8.850**
Average NTRs	8.68**

Table 4.12 summarises the data from the crosses between Cer10 and the recombinants taking into consideration the findings in TR 25_E11 and NTR 25_E01. Although removal from their respective groups does not have a large effect on the overall average hybrid necrosis rate of either group, due to the relatively small proportion of the total number of hybrids scored that originate from these two plants (~9%), the difference in the hybrid necrosis arising from these two plants is evident when their data is shown separately from their groups.

So, the large χ^2 value found when comparing the crosses within either the TR or NTR group can be attributed to two plants, TR 25_E11 and NTR 25_E01, whose rates of hybrid necrosis differ greatly from the other TRs and NTRs respectively. Clearly then, in these two plants recombination between the marker e242 and the *T* locus has disrupted the linkage between the *T* and *NEC* loci causing the tolerance-conferring allele (*T*) to be associated with the non-lethal allele of the *NEC* locus (*nec*) whilst the non-tolerant allele (*t*) is associated with the lethal allele *NEC.*

This switch in the linkage between copper tolerance and hybrid necrosis provides the first empirical evidence that the loci controlling these phenotypes are discrete and not synonymous.

Table 4.12 Summary of the average hybrid necrosis rates from crosses between Cer10 and recombinants. The data is shown both pooled for the TR and NTR groups and also with the data from TR 25_E11 and NTR 25_E01 omitted from their respective groups.

Cross (with Cer10)	Number of healthy hybrids	Number of necrotic hybrids	Total no. of hybrids scored	% Hybrid Necrosis
All TRs	1852	1778	3630	48.98
TRs with 25_E11 omitted	1569	1731	3300	52.45
25_E11	283	47	330	14.24
All NTRs	3514	371	3885	9.55
NTRs with 25_E01 omitted	3302	223	3525	6.33
25_E01	212	148	360	41.11

4.3.2 Identifying T locus markers and the recombination window containing NEC

Subsequent to the conclusion of the crosses between recombinants and Cer10 and analysis of the hybrid necrosis data they provided, additional markers linked to the *T* locus were identified and their genotypes in the TR and NTRs used to identify a recombination window containing the *NEC* locus.

The initial marker screening identified ninety polymorphic markers that varied in allele length between A and D. These markers were then tested in the progeny of a D x A cross and three were found to have genotypes strongly associated with the copper tolerance genotype. These are shown in table 4.13 along with a previously identified marker, e242.

As previously stated, the first marker to be identified, MgSTS242 (e242) was used as the marker to identify the recombinant plants used in the crosses to Cer10. A total of 3486 plants

were successfully genotyped at this marker and phenotyped for copper tolerance and recombination between it and the *T* locus was found to have occurred in sixty-eight plants.

Table 4.13 The identity and position of the four informative markers associated with the Copper tolerance locus and their genotypes in offspring of the D x A cross

Marker	**MgSTS242**	**Sg1788**	**e481/G12**	**Atgene5**
Location (scaffold position)	Scaf_177 505kb	Scaf_84 730kb	Scaf_84 539kb	Scaf_84 37kb
Tolerant genotype	168/178	335/334	368/0	235/237
TR genotype	178/178	334/334	0/0	237/237
Non-tolerant genotype	178/178	334/334	0/0	237/237
NTR genotype	168/178	335/334	368/0	235/237

The second marker identified was Sg1788 at 730kb on Scaffold_84. Normally, tolerant plants had the genotype *335/334* whereas non-tolerant plants had the genotype *334/334* meaning the *T* allele was associated with the 335bp allele at this locus. Where recombination had occurred between Sg1788 and the *T* locus, tolerant recombinants had the genotype *334/334* and non-tolerant recombinants had the genotype *335/334*. Out of the sixty-eight recombinants between e242 and *T*, forty-eight were also recombinants between Sg1788 and the *T* locus.

The third marker to be found was marker e481/G12 at 539kb on Scaffold_84. In progeny from the D x A cross, marker e481 had the genotype *368/0* in the copper tolerant offspring or the genotype *0/0* in the copper non-tolerant offspring. Thus, the 368bp allele is associated with the *T* allele and a null allele (*0*) is associated with the *t* allele. Accordingly, recombination between the *T* locus and e481 resulted in tolerant plants with the genotype *0/0* and non-tolerant plants with the genotype *368/0*. Out of the sixty-eight recombinants at e242, thirty-seven were recombinants between e481 and the *T* locus so that in these plants recombination occurs after the e481 locus.

The fourth marker identified was Atg5 at 37kb on Scaffold_84. Copper tolerant progeny from the D x A cross generally had the alleles 235/237 whilst non-tolerant progeny had just the 237 allele meaning the 235bp allele is usually associated with the *T* allele and the 237bp allele is usually associated with the *t* allele. Out of the sixty-eight e242/T locus recombinants plants successfully genotyped at this marker, twenty-seven were recombinants between Atg5 and the T locus.

The above data enabled the construction of a linkage map, shown in figure 4.2 which, combined with the two known recombinant plants between *T* and *NEC* (25_E11 and 25_E01), can be used to limit the location of the NEC locus.

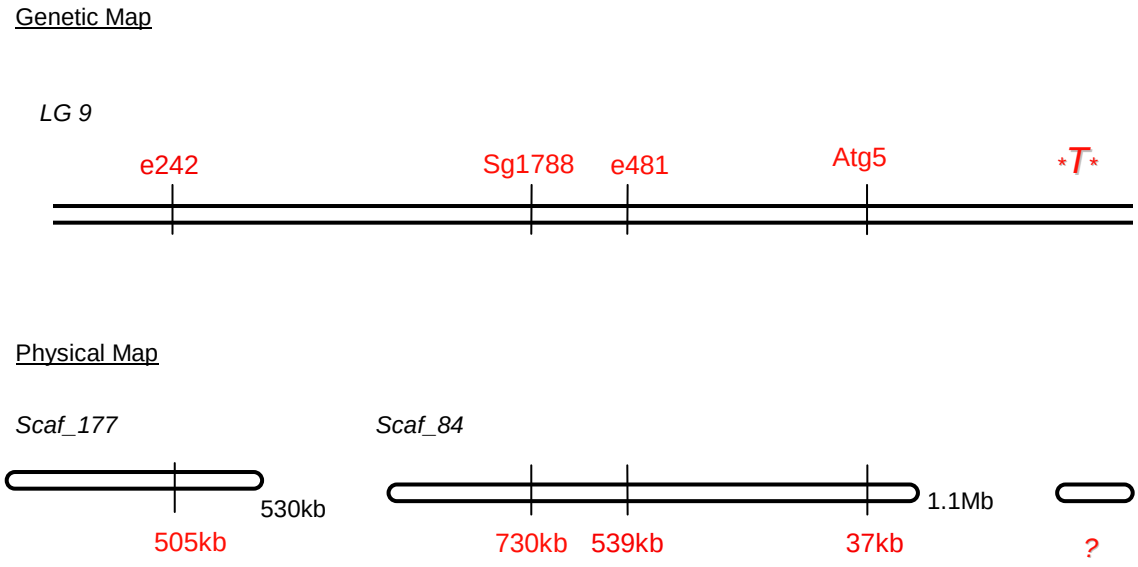


Figure 4.2 A genetic and physical map showing the positions of the four T locus markers identified by research reported in this chapter. The maps are constructed from Mimulus genome information available at www.mimulusevoution.org and marker mapping work carried out by the research team at Duke University. The order of markers on the genetic map was known prior to the crossing experiments, but their position relative to the T locus was established by this research.

In table 4.14, the marker genotypes in each of the TRs and NTRs crossed with Cer10 are shown in terms of whether it is the copper tolerant or non-tolerant genotype and thus shows the site of the recombination breakpoint in each plant. As the recombinant plants were identified by recombination between the *T* locus and e242, all the TRs are non-tolerant at e242 whilst all the NTRs are tolerant, but there is variation amongst the recombinants as to when the breakpoint occurs after e242. The informative genotypes in terms of identifying a window containing *NEC* are those of 25_E11 and 25_E01 when compared to the other TRs and NTRs with the same genotype.

In 25_E11, the recombination breakpoint is between the markers e481, which has a recombinant genotype, and Atg5, which has a non-recombinant genotype. This same genotype is found in seven other TRs yet the recombination event in these plants does not disrupt linkage between necrosis and copper tolerance. In theory, if the *NEC* locus was located between Sg1788 and e481 then the linkage between tolerance and necrosis would always be disrupted by a single recombination event occurring between e481 and Atg5 and so all seven of these TRs would have a phenotype similar to 25_E11. Conversely, if the *NEC* locus was located after Atg5 then linkage would not be disrupted in any plant by this recombination between e481 and

Atg5. As established in section 4.3.1, the linkage between tolerance and necrosis has been broken by the recombination event in 25_E11 but not in the other TRs. This variable disruption of the linkage between tolerance and necrosis by the position of the recombination breakpoint between e481 and Atg5 indicates that the *NEC* locus must lie in between markers e481 and Atg5 with the breakpoint usually occurring between e481 and *NEC* except in 25_E11 when it occurs in between *NEC* and Atg5. This implies the order of the loci is *e481 – NEC – Atg5 – T*.

Table 4.14 The marker genotypes and recombination breakpoints in the recombinants (between marker e242 and the T locus) that were crossed with Cer10 to establish the degree of hybrid necrosis caused. Informative plants are highlighted in bold. Plants with data inconsistent from other data are shown in italics.

[T = marker genotype usually associated with copper tolerance i.e. genotype in the copper tolerant parent plant D; NT = marker genotype usually associated with copper non-tolerance i.e. genotype in the non copper tolerant parent plant A]

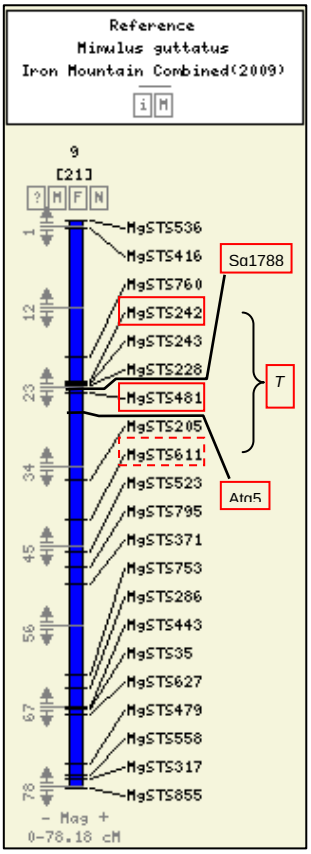
Plant	Hybrid necrosis (%)	Tolerance genotype at marker:				Copper Tolerance Phenotype
		e242	Sg1788	e481	Atg5	
			TR			
25_A09	*61.67*	*NT*	*NT*	*NT*	*NT*	*T*
28_G06	*44.58*	*NT*	*NT*	*NT*	*NT*	*T*
43_H01	*52.92*	*NT*	*NT*	*NT*	*NT*	*T*
25_E11	**14.72**	**NT**	**NT**	**NT**	**T**	**T**
26_C07	43.33	NT	NT	NT	T	T
26_H02	50.42	NT	NT	NT	T	T
32_G02	61.04	NT	NT	NT	T	T
34_A03	52.29	NT	NT	NT	T	T
35_F09	56.67	NT	NT	NT	T	T
43_G01	56.25	NT	NT	NT	T	T
53_B01	52.50	NT	NT	NT	T	T
			NTR			
10_F08	5.00	T	NT	NT	NT	NT
14_D05	4.58	T	NT	NT	NT	NT
19_C03	8.75	T	NT	NT	NT	NT
19_F06	6.33	T	NT	NT	NT	NT
29_G11	6.67	T	NT	NT	NT	NT
41_C05	6.00	T	NT	NT	NT	NT
43_B01	5.00	T	NT	NT	NT	NT
43_C01	4.17	T	NT	NT	NT	NT
46_B12	3.33	T	NT	NT	NT	NT
35_A12	7.22	T	NT	NT	NT	NT
25_E01	**41.11**	**T**	**T**	**T**	**NT**	**NT**
19_H12	**6.67**	**T**	**T**	**T**	**NT**	**NT**
24_D09	7.33	T	T	T	NT	NT
32_H09	6.39	T	T	T	NT	NT
35_B06	15.00	T	T	T	NT	NT
48_B01	7.08	T	T	T	NT	NT

However, the data from three TRs, 25_A09, 28_G06 and 43_H01, do not concur with this order of loci. In these three plants the recombination breakpoint lies between Atg5 and the *T* locus. If *NEC* is positioned between e481 and Atg5, recombination at this site would always break the linkage between the *NEC* and *T* loci and consequently all three of these TRs would have hybrid necrosis rates similar to 25_E11. However, all three cause a high degree of necrosis thus suggesting that the location of the *NEC* locus must be to the right of Atg5, either in between Atg5 and the T locus or to the right of the T locus so that the order is *e481 – Atg5 – NEC – T* or *e481 – Atg5 – T – NEC*.

Whilst the data from the other TRs that cause high necrosis can support this position for *NEC*, the recombination breakpoint and low necrosis data from 25_E11 oppose it. There are four possible explanations for this inconsistency in the data. First, 25_E11 could be a double recombinant in which the first recombination breakpoint occurs between e481 and Atg5 and the second breakpoint between the *NEC* and *T* loci so that the order along the chromosome is *e481 – Atg5 – NEC – T*. Second, the genotype at Atg5 might be scored with error. As the genotype of 25_E11 has been confirmed by re-genotyping at this marker, it is possible that the genotypes of 25_A09, 28_G06 and 43_H01 at Atg5 may be incorrect. Third, there may be an error in the tissue sampling or contamination of the DNA from 25_A09, 28_G06 and 43_H01 which would produce invalid marker genotypes. These two later explanations could be tested by repeating the sampling and genotyping of these plants. Fourth, 25_A09, 28_G06 and 43_H01 may be double recombinants, which is unlikely given the close proximity of the loci involved, or have experienced a gene conversion event.

The data from the crosses with the NTRs can provide some further evidence for the location of *NEC*. Each NTR has a breakpoint prior to Atg5, with the breakpoint in six plants, including 25_E01, occurring between e481 and Atg5. Whilst 25_E01 causes a high degree of necrosis, the other five NTRs with the same genotype e481 and Atg5 do not. As the disruption of the linkage between tolerance and necrosis is variable in these NTRs, these data concur with those of 25_E11 and the seven TRs that pinpoint the *NEC* locus as lying in between markers e481 and Atg5 so that the order is *e481 – NEC – Atg5 – T*. However, it is possible that double recombination could also have occurred in 25_E01.

As shown in figure 4.3, the EST markers e242 and e481 have been mapped to Chromosome 9. Marker Sg1788 and Atg5 are based on mRNA sequences on scaffold_84 near the e481 marker locus, and thus also map to this chromosome.



[From www.mimulusevolution.org]

Figure 4.3 A map of Chromosome 9 of the *M. guttatus* genome showing the location of the copper tolerance locus (T) and the four markers identified in this research as being linked to this locus and subsequently used to reveal the region containing the NEC locus. Although not used to identify the NEC locus region, marker MgSTS611 is highlighted in the map as it forms a probable limit to the region containing the T locus.

The recombination frequencies (RF), and thus the distance in centimorgans (cM), between the *T* locus and its linked markers can be calculated as described in 4.2.5 and these distances are shown on the map in figure 4.4. The RF between *NEC* and *T* was calculated using the formula *2/27 x 68/3486* = 0.14cM. In figure 4.4, this RF suggests that *NEC* should lie to the right of Atg5 which, with the exception of the three TRs 25_A09, 28_G06 and 43_H01, is considered inconsistent with the data in table 4.10. This is most likely due to the fact that only twenty-seven plants were scored for the necrosis phenotype, of which only two were recombinants between the *T* and *NEC* loci, whilst sixty-eight were scored for the markers, of which twenty-seven were recombinants between Atg5 and *T*. These ratios (2/27 and 27/68) are significantly different ($\chi 2$ = 5.745, significant at $p<0.05$, df = 1). In any event, RFs are not as accurate as the mapping data in pinpointing the region containing *NEC*.

The tight linkage between the loci in this region makes double recombination an unconvincing explanation. This, along with the data from 25_E11 and 25_E01 when compared to other TRs and NTRs with the same genotype both pointing towards the order of loci being *e481 – NEC – Atg5 – T*, suggests that an incorrect genotype at Atg5 for the three TRs (25_A09, 28_G06, 43_H01) is the most probable cause of the data inconsistency in the TR genotypes with regards to the location of *NEC*.

Thus, the combination of marker genotypes, copper tolerance phenotypes and hybrid necrosis phenotypes can be taken to reveal a 500kb recombination window between marker e481 at 539kb and marker Atg5 at 37kb on Scaffold_84 of the *Mimulus* genome assembly which contains the hybrid necrosis locus.

4.4 Discussion

This chapter presents the first empirical evidence that the *NEC* locus in Copperopolis is discrete from the locus conferring copper tolerance and identifies a 500kb region on Chromosome 9 of the *M. guttatus* genome as the location for this gene.

The conclusion that the copper tolerance and *NEC* loci are discrete is drawn from the analysis of hybrid necrosis data from crosses between a Cerig individual (Cer10) that possesses the complementary incompatible alleles which results in high hybrid necrosis when crossed to a plant possessing the *NEC* gene and both copper tolerant and non-tolerant plants in which recombination has occurred between the tolerance locus and a linked marker (e242).

In most crosses, tolerance is associated with high hybrid necrosis and non-tolerance is associated with low hybrid necrosis meaning that the recombination event has not disrupted the linkage between the *T* and *NEC* loci. However, in the progeny from crosses between Cer10 and one tolerant recombinant (25_E11) hybrid necrosis is low and between Cer10 and one non-tolerant recombinant (25_E01) hybrid necrosis is high. This indicates that, in these plants, recombination between e242 and the *T* locus has disrupted the association between tolerance and hybrid necrosis thus revealing that the loci controlling these phenotypes are discrete.

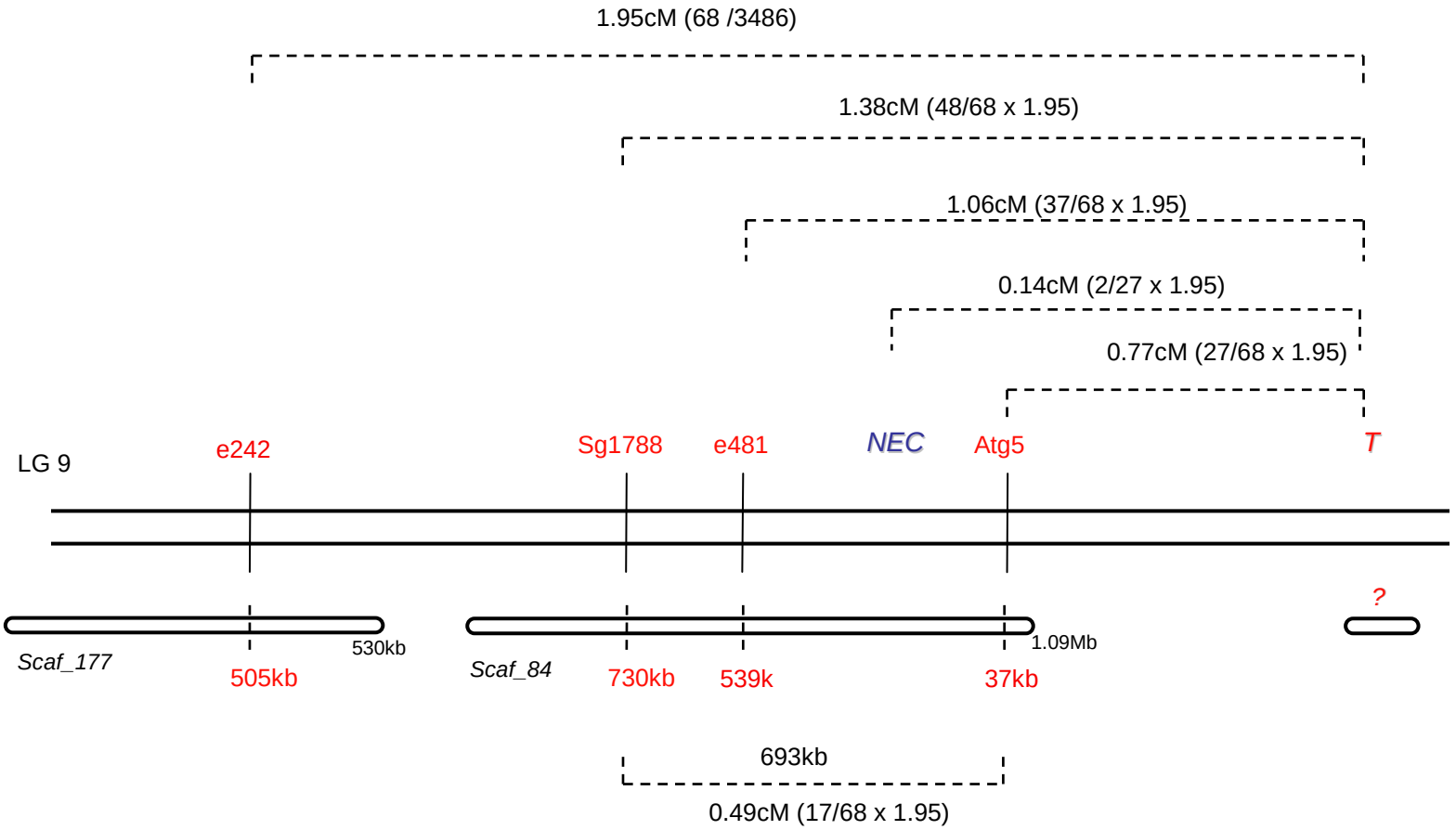


Figure 4.4 Map showing the positions and recombination frequencies/distance in cM between the T locus and the four linked markers and the NEC locus. Courtesy of Kevin Wright, Duke University.

Interestingly, there is heterogeneity in the degree of hybrid necrosis within the TR and NTR categories. In particular, the necrosis rate of 25_E11 is significantly higher than that of the NTRs whilst the necrosis rate of 25_E01 is significantly lower than that of the TRs. This could suggest that there is at least one other locus within the region of the hybrid necrosis locus that modifies the rate of necrosis.

Ideally, more recombinants would be used to conduct crosses as this would have enabled a more precise mapping of the interval containing the *NEC* locus. Due to genotyping failure as a result of noise or failed fragment amplification with PCR or ambiguity in the copper tolerance phenotype, some plants that were recombinant between the *T* locus and e242 were identified some time after crosses between recombinants and Cer10 had been performed and time and space constraints prevented the possibility of including these individuals in the test crosses.

The ambiguity in the copper tolerance phenotypes of some D x A plants during the phenotyping for identification of recombinants was due to failure of the cutting to produce either roots or blackened root primordia or to the cutting having too short a stem to be phenotyped even though other plants in the tray were tall enough, perhaps as a result of crowding. Had this not occurred there would have been a greater overall sample size and so it might have been possible to identify more recombinant plants as a whole, thereby providing a greater chance of finding recombinants between the *T* and *NEC* loci. The strength of the linkage between *T* and *NEC* implies that recombination between them is not that common and so the larger the sample size the more chance there would be of identifying plants in which tolerance and hybrid necrosis had recombined.

In order to assuage the fact that there were only two recombinants with the break in the *T* and *NEC* linkage and to reinforce the data from them, these two plants were phenotyped multiple times to ensure the copper tolerance phenotype was correct: TR 25_E11 was phenotyped four times and NTR 25_E01 was phenotyped five times. In addition, six replicate crosses to Cer10 were performed with each of these recombinants to ensure that the switch in necrosis rates was consistent and therefore not due either to error in scoring hybrid necrosis or the effect of the growing environment causing death through factors other than hybrid necrosis. In spite of the low sample size, the data from these two plants are consistent and, due to the very large difference they exhibit, provide compelling evidence that the *T* and *NEC* loci are discrete.

The copper tolerance locus and the four linked markers identified in this chapter of research have been mapped to Chromosome 9. The first marker identified, e242, was the marker used to identify all the recombinant plants used in the crosses with Cer10. Subsequent to the completion of crosses between Cer10 and the e242/*T* locus recombinants, three additional markers for the *T* locus were found. Although they were not used to identify recombinant plants

for use in crosses, these markers were critical in identifying the region containing the *NEC* locus.

Collectively, the recombination breakpoints revealed by the marker genotypes and the copper tolerance phenotypes of the TR and NTR plants reveal a 500kb recombination window containing the Copperopolis *NEC* gene. This gene is incompatible with alleles of a small number of genes in the Cerig population (Macnair & Christie, 1983) and, when brought together in the genome of the hybrid, cause necrosis, dwarfism and ultimately death.

4.5 Conclusion

A single gene in the Copperopolis population of *M. guttatus* interacts with the small number of polymorphic genes in the Cerig population to cause hybrid necrosis via Dobzhansky-Muller incompatibilities (Macnair & Christie, 1983). This gene was demonstrated by Macnair & Christie (1983) to be either synonymous with, or very tightly linked to, the single major locus conferring copper tolerance (Macnair, 1981, 1983).

This chapter has revealed that the locus for hybrid necrosis (*NEC*) and the locus for copper tolerance (*T*) are discrete. Consequently, identifying the *NEC* gene will yield interesting information on the function of this gene and the possible forces that caused the evolution of hybrid necrosis between these two populations. Fine-mapping of the 500kb region identified in this chapter as containing *NEC* and subsequent bioinformatic analysis on this region will facilitate the identification of a limited number of candidate genes for *NEC*.

Chapter 5

Bioinformatics and fine-mapping of the recombination interval containing the hybrid necrosis (*NEC*) locus in Copperopolis

5.1 Introduction

Hybrid necrosis, an intrinsic postzygotic reproductive isolating barrier, in the progeny of crosses between the copper tolerant 'Copperopolis' (Cop) and copper non-tolerant 'Cerig' populations of *Mimulus guttatus* results from a Dobzhansky-Muller (DM) genic incompatibility (Macnair and Christie, 1983). The DM model posits that hybrid incompatibilities are the result of deleterious epistatic interactions between incompatible alleles of complementary genes which have functionally diverged at a minimum of two loci resulting in the formation of new alleles and that at least one allele from each species must be involved (Coyne and Orr, 2004, Dobzhansky, 1937, Muller, 1942). In Chapter 4, a 500kb region in chromosome 9 of the *M. guttatus* genome was identified as the location of the single incompatibility locus, *NEC*, from the Cop population that interacts with genes in the Cerig population to cause hybrid necrosis.

Studies of hybrid dysfunctions arising from such genic incompatibilities have thus far identified a number of genes that are involved in DM interactions (see Chapter 1). Although these genes cause lethality or sterility in hybrid progeny, those that have been the subject of more detailed molecular analysis have also been found to perform normal biological functions within their own species. These functions, summarised in table 5.1, are diverse and range from gene regulation to enzyme activity to structural proteins. It is evident, then, that DM genes do not exist solely as barriers to reproduction but that they also fulfil important functional roles within a species. Consequently, this makes it challenging from a bioinformatics perspective to identify candidates for a DM gene. Indeed, all annotated genes are potential *NEC* candidates and thus it is crucial to narrow down the size of the genomic region identified as containing *NEC* to limit the number of possible candidate genes.

Table 5.1 A summary of Dobzhansky-Muller incompatible genes whose biological functions have been identified

Species	Gene	Function	Hybrid…	References
Drosophila simulans	*Lhr*	Heterochromatin binding	Lethality	Watanabe, 1979; Nolte et al, 2008; Barbash et al, 2000, 2003; Brideau et al, 2006
Drosophila melanogaster	*Hmr*	DNA-binding/ Gene regulation	Lethality	Hutter & Ashburner, 1987; Hutter et al, 1990; Orr et al, 1997; Orr & Irving, 2000; Barbash et al, 2000, 2003; Brideau et al, 2006; Bolkan et al, 2007
Drosophila melanogaster	*Zhr*	Possible role in chromosome segregation in mitosis	Lethality	Sawamura et al 1993c; Ferree & Barbash, 2009
Drosophila simulans	*Nup 96*	Nucleoporin	Lethality	Presgraves et al, 2003; Tang & Presgraves, 2009
Drosophila simulans	*Nup160*	Nucleoporin	Lethality	Tang & Presgraves, 2009
Xiphophorus maculates	*Xmark2*	Tyrosine kinase	Lethality	Wittbrodt et al, 1989; Wu & Ting, 2004; Schartl, 2008
Lycopersicon pimpinellifolium	*Cf-2*	*R* gene	Necrosis	Kruger et al, 2002
Lycopersicon esculentum	*RCR3*	Interacts with *R* gene products	Necrosis	Kruger et al, 2002
Lycopersicon	*Cf-9*	*R* gene	Necrosis	Wulff et al, 2004
Arabidopsis thaliana Uk-1	*DM2*	*R* gene	Necrosis	Bomblies et al, 2007
Arabidopsis thaliana Uk-3	*DM1*	*R* gene	Necrosis	Bomblies et al, 2007
Lactuca saligna	*RIN4*	Interacts with *R* genes	Necrosis	Jeuken et al, 2009
Oryza sativa L. cv. *indica*	*hbd2*	Casein kinase 1	Necrosis	Matsubara et al, 2007; Yamamoto et al, 2007, 2010
Oryza sativa L. cv. *japonica*	*hbd3*	*R* gene	Necrosis	Matsubara et al, 2007; Yamamoto et al, 2007, 2010
Oryza sativa L. cv. *indica*	*hwh1*	Glucose-methanol-choline oxidoreductase	Necrosis	Jiang W. et al, 2008
Oryza sativa L. cv. *japonica*	*hwh2*	Hexose transporter	Necrosis	Jiang W. et al, 2008
Oryza sativa L. cv. *indica*	*hwe1*	Possible rice blast resistance gene	Necrosis	Kubo & Yoshimura, 2002
Drosophila mauritiana	*OdsH*	Transcription factor	Sterility	Ting et al, 1998
Drosophila pseudoobscura pseudoobscura	*Ovd*	DNA-binding	Sterility	Phadnis & Orr, 2009
Mus musculus	*Prdm9*	Histone H3 methyltransferase	Sterility	Mihola et al, 2009
Saccharomyces cerevisiae	*MRS1*	Splices mitochondrial *COX1* introns	Sterility	Chou et al, 2010
Saccharomyces cerevisiae	*AIM22*	Ligase for mitochondrial protein lipolyation	Sterility	Chou et al, 2010
Saccharomyces bayanus	*AEP2*	Mitochondrial translation protein	Sterility	Lee et al, 2008
Saccharomyces cerevisiae	*OLI1*	ATP synthase	Sterility	Lee et al, 2008

Hybrid necrosis is a common isolating barrier in many different species of plants and shares characteristic phenotypes with those seen in plants responding to pathogens and consequently has implicated the involvement of plant immune system components in hybrid necrosis (Bomblies and Weigel, 2007, Ispolatov and Doebeli, 2009). Plant disease resistance genes (*R* genes) have a crucial role in triggering the hypersensitive response (HR) leading to the programmed cell death (PCD) associated with a plants' response to pathogens.

Autonecrosis, a syndrome that is milder than hybrid necrosis but shares similar characteristics, has been attributed to genes known to interact with *R* genes (Kruger et al., 2002, Wulff et al., 2004). Moreover, several cases of hybrid necrosis involving *NB-LRR* genes, the largest class of *R* genes, have been identified (Alcazar et al., 2009, Bomblies et al., 2007, Yamamoto et al., 2010). As such, aberrant interactions involving *R* genes can potentially cause autoimmune-like incompatibility resulting in PCD and forming a range of responses in the hybrids, from disease resistance to severe necrosis or lethality (Bomblies and Weigel, 2007).

Collectively, these studies reveal compelling evidence for the involvement of plant immune system genes, particularly *R* genes, in hybrid necrosis. Given the similarity in the phenotypes of hybrid necrosis in the literature and in this project, there is a sound argument for supposing *R* gene involvement in the DM incompatibility interaction between Cop and Cerig. However, despite the increasing number of examples of hybrid dysfunctions resulting from DM incompatibilities, relatively few DM genes have been identified and fewer still have a known biological function and so it remains a significant prospect that genes with a function other than in disease resistance can interact to cause hybrid necrosis.

Currently, the only hybrid necrosis genes with known function are involved in disease resistance pathways. Apart from *R* genes, which are involved in at least five cases of hybrid necrosis, other hybrid necrosis genes whose functions have also been identified include *RIN4*, which encodes a protein that interacts with *R* genes products (Jeuken et al., 2009), *Hbd2*, which encodes a casein kinase that interacts with an *R* gene (Yamamoto et al., 2010), and *hwh1* and *hwh2* which encode an oxidoreductase and a hexose transporter respectively ((Jiang W et al., 2008) both of which may also be involved in the disease response (Herbers et al., 1996, Lamb and Dixon, 1997). Although other genes causing necrosis have been identified (see Chapter 1, table 1.3), none have yet had their biological function determined. In light of this, identifying the hybrid necrosis gene in Cop would provide valuable insight into whether it is another example of a gene involved in the immune response or whether it is a gene with an entirely different function that is capable of causing hybrid necrosis in plants.

In Chapter 4, a 500kb region between two markers, STSe481 and Atg5, on chromosome 9 of the *M. guttatus* genome was identified as containing the *NEC* locus. The aim of this next section of research is to establish the presence or absence of *R* genes within the 500kb window and to identify all candidate genes for the *NEC* locus. This is achievable by using a combination of tools. Molecular techniques can be used to develop additional markers to enable fine-mapping and refinement of the 500kb window and bioinformatics can be applied to annotate genes within the window and predict peptide sequences enabling putative conserved domains to be identified, and possible gene function to be predicted and the likelihood of a candidate gene being the *NEC* gene to be assessed.

5.2 Methods

Full details on all the plants, methods and techniques used in this research are provided in Chapter 2.

The on-going *Mimulus guttatus* genome sequence on the GBrowse page at http://www.mimulusevolution.org and also at http://monkeyzome.phytozome.net/cgi-bin/gbrowse/mimulusv11/ was used to view the *M. guttatus* sequence assembly and preliminary annotations. Included in the annotation, and used in this project, are sequences for mRNA transcripts/predicted *M. guttatus* genes as well as details on markers, BAC sequences and *Arabidopsis thaliana* protein homologs. In January 2010, Phytozome version 5.0 (http://www.phytozome.net/) became available, and incorporates both updated and newly released plant genomes, one of which is *Mimulus guttatus*. This site provides access to twenty sequenced and annotated plant genomes and where possible, each gene has been annotated with protein family database (PFAM), Eukaryotic Orthologous Groups (KOG), Kyoto Encyclopaedia of genes and genomes (KEGG), and Protein Analysis Through Evolutionary Relationships (PANTHER) assignments.

5.2.1 Fine-mapping of the recombination window containing NEC

The region to be fine-mapped was a 500kb window on scaffold_84 of the 7x *Mimulus* genome assembly between markers e481 at 540kb and Atg5 at 37kb which was identified in Chapter 4 as containing the *NEC* locus.

5.2.1.1 Provenance of plants

The plants used in the fine-mapping of the window containing the *NEC* locus were those used or identified in Chapter 4. In brief, markers were initially screened for intron length polymorphisms between the copper tolerant parent plant D and the copper non-tolerant parent plant A. Those that showed variation were then tested in selected offspring of a D x A cross, primarily the copper tolerant recombinant TR 25_E11 and the non-tolerant recombinant NTR 25_E01 in which the recombination between the *T* locus and a linked marker (e242) disrupted the linkage between the copper tolerance (*T*) locus and the *NEC* locus thereby identifying them as distinct loci. As a control, the markers were also either tested in TR 32_G02 and NTR 19_H12, both recombinants (copper tolerant and non-tolerant respectively) between the *T* locus and the marker e242 but in which the linkage between *T* and *NEC* had not been disrupted by the recombination event, or in the copper tolerant plant 46_B06 and the non-tolerant plant 46_B08, both of which were non-recombinant.

5.2.1.2 Primer design for new markers

The first round of marker screening was for intron length polymorphisms in markers based on mRNA (JGI annotation) sequences. The FASTA sequences of nine mRNAs within the 500kb window on scaffold_84 were extracted from GBrowse (www.mimulusevolution.org) and primers were designed in exons flanking the introns. In order to maximise the likelihood of finding a polymorphic marker, primers were designed for every suitable intron (more than 150bp and flanked by exons/CDS of at least 20bp in which to design the primers) so that more than one marker was designed for some of the mRNAs. This gave a total of fifteen new markers spaced between 96kb and 539kb throughout the region to be fine-mapped. Their relative positions along scaffold_84 are shown in figure 5.1.

Ideally, all the new markers should have been evenly spaced at 50-70kb intervals but this was not possible as some mRNA sequences could not be used to design primers as they had either no introns, introns that were too short (<150bp), or exons that were too short to use for primer design (less than 20bp). Consequently some markers were closer together than 50kb: for example, there were four new markers to be screened within a 30kb area around marker e481. However, having markers overlapping or in close proximity to each other can be beneficial as it enhances the likelihood of finding at least one that is polymorphic.

The mRNA FASTA sequences were put into an online programme, Primer3 http://frodo.wi.mit.edu/primer3/ , which designed pairs of primers. Each forward primer had an additional sequence, ‘M13’, attached to the 5’ end. This fluorescently-labelled M13 primer

sequence, which works for all primers sets, can be read by the capillary sequencer on which the PCR products were sequenced and thereby enable sequencing without incurring the cost associated with labelling each primer individually (see methods 2.6.2).

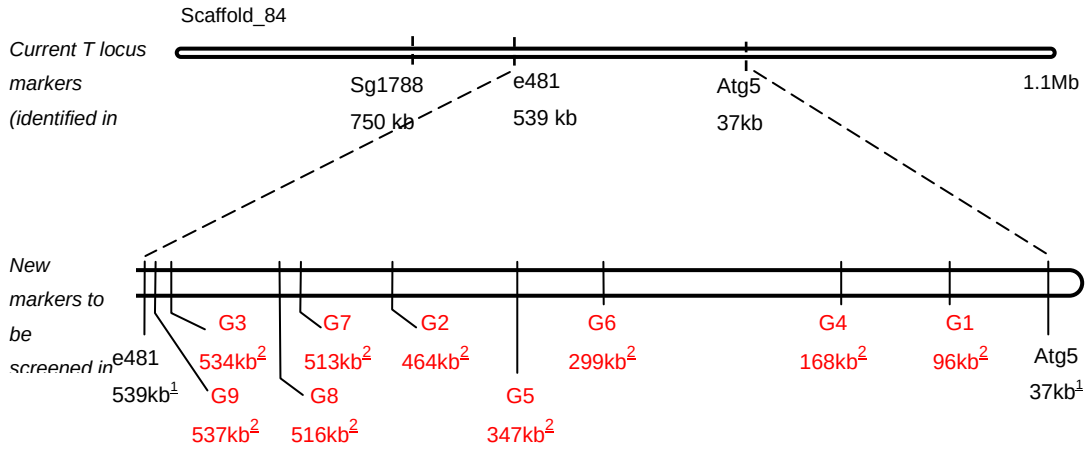
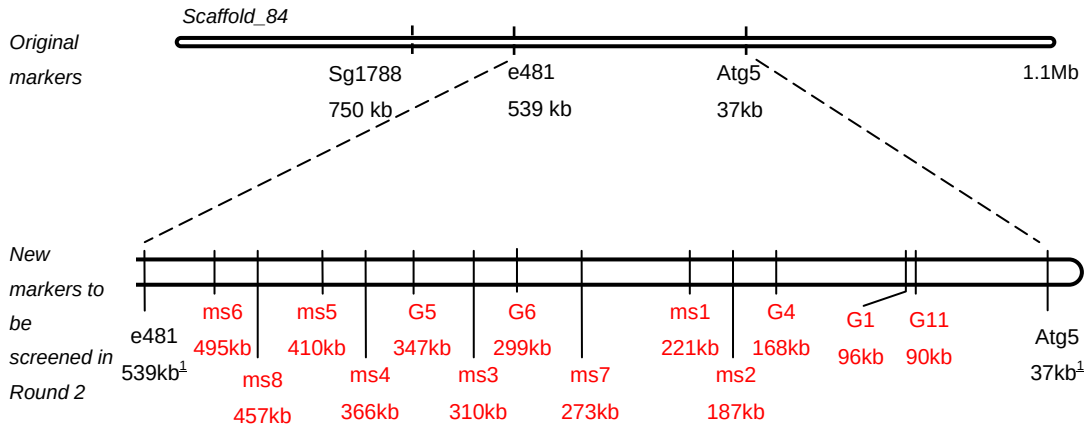


Figure 5.1 The positions along scaffold_84 of the Mimulus genome sequence of the nine mRNAs whose sequences were used to design the fifteen markers screened for variation in plants A (copper non-tolerant) and D (copper tolerant) in the first round of screening.

The second round of marker screening used five mRNA-based markers, the primers of which were designed using the method described above, and eight microsatellite markers. Primers for the microsatellite markers were developed by inputting the entire scaffold_84 sequence into the programme Perfect Microsatellite Repeat Finder (http://sgdp.iop.kcl.ac.uk/nikammar/repeatfinder.html) and selecting the sequences that were of good size (200- 600bp) and evenly spaced within the 100-500kb. The relative positions of the second set of markers along scaffold_84 are shown in Figure 5.2.



[1] confirmed marker; [2] new marker to be tested

Figure 5.2 The positions along scaffold_84 of the Mimulus genome sequence of the five mRNA markers and eight microsatellite markers screened for variation in plants A and D in the second round of screening. The third and final round of marker screening was for markers based on single nucleotide polymorphisms (SNPs) within three genes in the window defined by two new markers identified during the first two rounds of screening. Using the Primer3 programme, primers were designed in introns flanking the largest exon for each of the genes and used to PCR-amplify the exon sequence. The genes on which these SNP markers were based were positioned at 180kb, 234kb and 297kb on scaffold_84.

5.2.1.3 Identification of informative markers

DNA was extracted from the parent plants A and D as well as the selected recombinants and non-recombinants using the CTAB method and used in PCR reactions (see 2.6.3 “M13 primers”) with the marker primers described above. The PCR programme used for all these reactions is given in 2.6.3.2.

PCR reactions were checked for successful amplification by gel electrophoresis. Those PCRs that were not successful were repeated with DNA from a second extraction and/or altered PCR parameters, or, if three attempts at amplification failed, alternative primers were redesigned if the marker was considered important in terms of its location i.e. where there were no other markers in close proximity.

Screening for new markers in the selected region containing *NEC* was based on possible intron length polymorphisms (rounds one and two) or microsatellites (round two) depending on whether they were in the A (copper non-tolerant) or the D (copper tolerant) parent plants.

Successful PCR reactions were sent for genotyping which was performed by subjecting the PCR-amplified DNA fragments with the incorporated 5' fluorescent-labelled primer (M13) to capillary electrophoresis and fragment analysis on an ABI 3730x1 DNA Analyzer (Applied Biosystems). The size of the amplified fragments was scored automatically by the programme GeneMarker (SoftGenetics, 2005, State College, PA) and the score verified by eye.

Screening in round three was based on SNPs in the largest exon of selected genes in the region containing *NEC*. The PCR product was sequenced using a BigDye sequencing reaction and SNPs were identified by eye. Those markers that were polymorphic in the parents were subsequently tested in selected recombinants and non-recombinants using the same method described above.

The initial results of marker screening are reported here partly to aid clarity in the method description and partly as the focus in the results section is on informative markers only.

Table 5.2 shows the results of the first round of marker screening in A and D. Successful amplification with DNA from both A and D was achieved with nine of the fifteen marker primers and the fragment analysis data on intron length (i.e. genotype) is shown. Of these nine markers, eight had variation in their intron length depending on whether they were in the A or the D parent. Four primers gave no fragment amplification and two only had amplification with DNA from one parent. PCRs with these primers were repeated three times in an attempt to achieve amplification but did not amplify a product and so ultimately these primers were abandoned as, with the exception of G1.I1 and G2.I1, they had other polymorphic markers in close proximity.

Of these polymorphic eight markers, four (G5.I1, G5.I4, G6.I4 and G9.I5) were then tested in four D x A progeny that had been identified as recombinants by the work in Chapter 4. These were the two recombinants that showed the break in linkage between the *NEC* locus and the copper tolerance locus (25_E11 and 25_E01), and two that did not show a break in the linkage - a TR that produced the expected high hybrid necrosis in crosses with Cer10 (32_G02) and a NTR that produced the expected low hybrid necrosis in crosses with Cer10 (19_H12). The genotypes are given in the results section. Of the remaining four polymorphic markers, G4.I1 was not tested until the second round of screening for practical reasons regarding cost efficiency whilst the other three (G3.I1, G7.I1 and G9.I8) were not tested because they either gave weak signals or had a large amount of stutter/noise making it difficult to determine the genotype and were all non-essential as they were very close to, or overlapped with, other markers that were also polymorphic.

Table 5.2 The fragment sizes (genotypes) in the parent plant lines A and D of the fifteen markers genotyped in the first round of marker screening.

*(italics = polymorphic; * = tested in selected D x A progeny)*

Marker	JGI ID (mRNA), position on scaffold_84 (kb), intron (I) number	Fragment size (bp) in plant 'A'	Fragment size (bp) in plant 'D'
G1.I1	mgf017327, 96278-98113, I2	No amplification	No amplification
G2.I1	mgf004251, 464335-471686, I1	No amplification	No amplification
G2.I3	mgf004251, 464335-471686, I3	524	524
G3.I1	*mgf022549, 534411-534953, I1*	*228*	*229*
*G4.I1**	*mgf019257, 167943-169726, I1*	*296/297*	*285/286*
*G5.I1**	*mgf001132, 346988-352861, I1*	*513*	*503*
G5.I3	mgf001132, 346988-352861, I3	No amplification	No amplification
*G5.I4**	*mgf001132, 346988-352861, I4*	*404/428*	*404*
*G6.I4**	*mgf001028, 298632-301303, I4*	*468*	*450*
G7.I1	*mgf018847, 513093-515793, I1*	*438*	*356*
G8.I1	mgf007155, 515931-522979, I1	No amplification	No amplification
G8.I6	mgf007155, 515931-522979, I6	881	No amplification
G9.I3	mgf003334, 537003-543715, I3	532/565	No amplification
*G9.I5**	*mgf003334, 537003-543715, I5*	*368/372*	*372/424*
G9.I8	*mgf003334, 537003-543715, I8*	*250/251*	*214/251*

At this time, a second round of screening was performed using a new set of primers. These included a redesigned primer for G1 as it failed to amplify a fragment in the first round of screening and had no other successful markers in close proximity, primers for a new mRNA marker (G11) at 90kb as an alternative if the new G1 primers failed again, and redesigned primers for markers at important positions (G4, G5 and G6) in case the markers identified in the first screen did not give informative results when tested in the D x A recombinants. In addition, to facilitate identifying as many markers as possible eight new markers were designed, this time based on microsatellites (see Figure 5.2).

The five new mRNA-based primers and two of the microsatellites, ms1_221k and ms2_187k, were used in the second round of marker screening performed using the same method as in the first round. PCR fragment amplification with DNA from both parents A and D was achieved with five of these eight primers (G4.2ndGen, G5. 2ndGen, G11.2ndGen, ms1_221k and ms2_187k) and the fragments were sent for genotyping using an ABI 3730x1 DNA Analyzer and GeneMarker.

Table 5.3 shows the results for the second round of marker screening in A and D. Of the seven new markers, G4.2ndGen, G5.2ndGen and ms1_221k showed variation in A and D. These were then tested in the recombinants TR 25_E11 and NTR 25_E01. However, instead of being tested

in the two other recombinants, as in round one, the markers were tested in two progeny from the D x A cross that were non-recombinants between the copper tolerance locus and its linked marker e242. One individual was copper tolerant (46_B06) and one was non-tolerant (46_B08). This was a stronger control than the use of the two recombinants in which the tolerance and hybrid necrosis linkage was not disrupted (32_G02 and 19_H01) as it enabled the determination of the marker genotypes in non-recombinant progeny in which it was expected that the fragment lengths of the marker would remain linked to the original parental phenotype of tolerance/non-tolerance thereby enabling easier identification of which allele was linked to the copper tolerance allele (*T*) and which to non-tolerance (*t*). The remaining six other microsatellite markers were not screened as success in identifying other markers at this point (see results section) rendered these markers of no value as they were all outside the recombination window identified by the new markers.

Table 5.3 The fragment sizes (genotypes) in the parent plant lines A and D of the seven markers genotyped in the second round of marker screening.

*(italics = polymorphic; * = tested in selected D x A progeny)*

Marker Name	JGI id (mRNA), position on scaffold_84)	Fragment size (bp) in plant 'A'	Fragment size (bp) in plant 'D'
G1.2ndGen	mgf017327, 96278-98113	No amplification	765
*G4.2ndGen**	*mgf019257, 167943-169726*	*631*	*651*
*G5.2ndGen**	*mgf001132, 346988-352861*	*797*	*780/790*
G6.2ndGen	mgf001028, 298632-301303	No amplification	211/360/543
G11. 2ndGen	mgf017956, 89957-91466	223	223
*ms1_221k**	*221 kb*	*204/270/274/278*	*200/204*
ms2_187k	187 kb	219/222/225	219/222/225

At the end of the second round of screening, of the twenty-two new markers screened a total of ten intron-length markers and one microsatellite marker showed polymorphism between plants A and D and therefore linkage to the copper tolerance phenotype. Seven intron-length markers and one microsatellite marker were genotyped in selected D x A progeny with the aim of using the genotype in conjunction with the plants' copper tolerance phenotype to establish whether or not recombination had occurred between the marker and the tolerance locus and thereby refine the recombination window to under 200kb. The results are described in the section 5.3.1.

Subsequent to an initial bioinformatics analysis, it was evident that the recombination window required further refining in order to reduce the number of candidate genes for *NEC*. To achieve this, three additional markers based on SNPs within gene exons were tested in a third and final

round of marker screening. Using Primer3, primers were designed flanking the largest exon for each selected gene (at 297kb, 234kb and 180kb). A BigDye sequencing reaction was performed directly from the PCR product and SNPs were verified by eye. SNPs were identified in each exon and are described in the results section.

5.2.2 Bioinformatics

After marker screening, the refined recombination window containing the hybrid necrosis locus (*NEC*) was viewed using the GBrowse function on two different websites: www.mimulusevolution.org and www.monkeyzome.phytozome.netv1.1 to identify all mRNA transcripts contained within it. Both websites were used for purposes of verification. The mRNA sequences were converted into their predicted peptide sequences using GENSCAN and then run through the NCBI BlastP to check for putative conserved domains. Both websites reported the same mRNAs with the same predicted peptide sequence and putative conserved domains. Along with NCBI, the functional annotation resources supplied by Pfam (http://pfam.sanger.ac.uk/), PANTHER (http://www.pantherdb.org/), and EMBL (http://www.ebi.ac.uk/embl/) were also used in the analysis of the putative conserved domain of each mRNA.

The *Arabidopsis thaliana* protein homologs within the window containing *NEC* were viewed on the www.mimulusevolution.org GBrowse site and their biological function details were found using the Arabidopsis Information Resource (TAIR: http://www.arabidopsis.org/). Additional information on the functions of the classes of protein homologs was found using internet and literature searches and further evidence in the literature was sought for whether any of these gene types had been previously implicated in other hybrid lethality studies.

Once the more comprehensive Phytozome version 5.0 (http://www.phytozome.net/) site was available (January 2010), it was used to verify the previous findings in terms of the functional annotation of the relevant region of the *Mimulus guttatus* genome.

5.3 Results

5.3.1 Fine mapping

Of the twenty-two markers screened for intron length/microsatellite polymorphisms, eleven were identified with polymorphisms linked to copper tolerance. Of these, eight were tested in selected D x A progeny, and three gave informative genotypes as shown in table 5.4.
The other five polymorphic markers tested in the selected D x A progeny (G5.I1, G5.I4, G6.I4, G4.2ndGen and ms1_221k) had genotypes that were either not informative or were inconclusive and so are omitted from table 5.4. Data from G5.I1 was not usable as the large quantity of stutter (presumably due to repetitive sequence or long homopolymer tracts) meant ascertaining an accurate genotype was not possible, G5.I4 showed no variation amongst the tolerant and non-tolerant plants (all had the genotype *404/428*) whilst G4.2ndGen had no heterozygotes in the recombinants/non- recombinants. Markers G6.14 and ms1_221k gave inconclusive data as when the marker was re-tested the allele signals did not consistently show up. The three polymorphic markers that were not tested in D x A progeny (G3.I1, G7.I1 and G9.I8) were excluded because they were in very close proximity (within 3kb) to other polymorphic markers and so the marker that gave the strongest or clearest signal in the sequencer was preferred for testing in the selected D x A progeny.

Table 5.4 The genotypes of the three informative markers that are linked to copper tolerance

*[x = marker not tested; * = only informative allele reported; T = copper tolerant; NT = non-tolerant; R = recombinant between marker e242 and T locus]*

		Genotype (allele size in bp) in plant…							
Marker	**scaffold_84 position**	**D (T)**	**A (NT)**	**25_E11 (TR)**	**25_E01 (NTR)**	**32_G02 (TR)**	**19_H12 (NTR)**	**46_B06 (T)**	**46_B08 (NT)**
G9.15	537kb	*372/424*	*368/372*	*372**	*424**	*372**	*424**	x	x
G5.2nd Gen	347kb	*780/790*	*797*	*780/797*	*790/797*	x	x	*790/797*	*780/797*
G4.I1	168kb	*285/286*	*296/297*	*286/297*	*285/297*	x	x	*286/297*	*285/297*

Marker G9.I5 showed variation linked to copper tolerance with a *424*bp allele associated with the tolerant (*T*) allele and a *372*bp or *368*bp allele associated with the non-tolerant (*t*) allele. The *T*-linked *424*bp allele was absent in the two tolerant recombinants (TR) but present in the two

non-tolerant recombinants (NTR) thus indicating recombination between the *T* locus and this marker. As this marker is positioned at 537-543kb it overlaps with marker e481 and so provides positive confirmation for this marker.

The genotype of the tolerant parent plant (D) at marker G5.2ndGen was *780/790* with the *790*bp allele being linked to the *T* allele whilst the *780*bp allele associated with the *t* allele (deduced from the genotypes of the two D x A offspring in which recombination has not occurred between e242 and the *T* locus: 46_B06 and 46_B08). The marker genotype of the non-tolerant parent (A) was *797*bp. The two non-recombinants both possess the allele associated with tolerance or non-tolerance that would be expected given their copper phenotypes. However, when tested in TR 25_E11 the *780*bp allele, usually associated with non-tolerance, occurs whilst in the NTR 25_E01 the *790*bp allele, usually associated with tolerance, occurs. This means that a recombination breakpoint is sited between this marker and the *T* locus, and as the linkage between *NEC* and *T* is disrupted by this recombination event in the two recombinant plants tested, the *NEC* locus must lie in between G5.2nd and the *T* locus. As such, this marker can replace e481 and redefine one limit of the recombination window containing *NEC* at 347kb.

At marker G4.I1, the genotype of the tolerant parent was *285/286* where, based on the genotype and phenotype of the non-recombinants, the *286*bp allele was associated with the *T* allele and the *285*bp allele was associated with the *t* allele, whilst the genotype associated with non-tolerance was *296/297*. At this marker, the tolerant recombinant and tolerant non-recombinant plants have the same genotype of *286/297* which is a tolerant genotype, and similarly the non-tolerant recombinant and non-tolerant non-recombinant plants have the same genotype of *285/297* which is a non-tolerant genotype. As the linkage between the marker alleles and tolerance has not been disrupted, these data indicate that the recombination breakpoint is prior to this marker. As such, this marker therefore indicates a limit to the breakpoint region, previously defined by the marker Atg5 at 37kb, and establishes it at 168kb.

So, fine-mapping revealed three additional markers for the recombination window defined in Chapter 4. Shown in figure 5.5, these markers are G9.I5 at 537-543kb, G5.2ndGen at 347-353kb and G4.I1 at 168-170kb. As G9.I5 overlaps with e481 (539-541kb), the key conclusion from this marker screening is that two of these markers, G5.2nd Gen and G4.I1, have reduced the region in which the *NEC* gene is located to a 179kb window between 347kb and 168kb on scaffold_84.

Bioinformatic analysis on this 179kb window showed that there were eight mRNAs encoding predicted peptides with putative conserved domains belonging to seven different protein superfamilies meaning further refinement was necessary in order to establish a more limited number of candidate genes for the *NEC* gene. In order to further refine the recombination

window, direct sequencing of three candidate genes was performed to detect single nucleotide polymorphisms (SNPs).

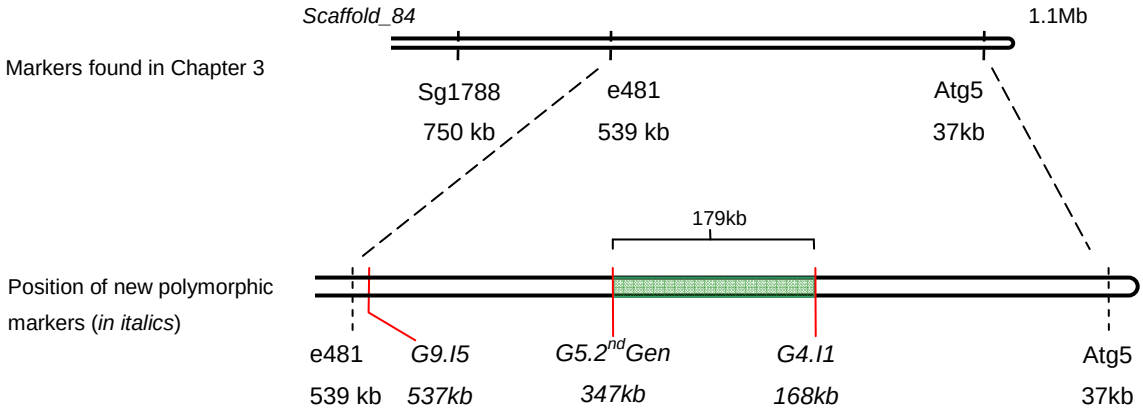


Figure 5.3 The positions along scaffold_84 of the Mimulus genome of the informative copper-tolerance linked markers that identify the window containing the NEC locus (highlighted)

The three candidate genes screened for SNPs were approximately 50kb apart within the 179kb window. These were genes at 297kb, 234kb and 180kb. As shown in table 5.5, these SNP markers each showed polymorphism at multiple bases within the A and D lines and the selected D x A offspring. These markers were used to define a smaller recombination window as described below.

Table 5.5 The genotype (nucleotide) at different bases of the three SNP markers screened for variation in plants A and D and selected offspring.

Plant	Marker Pr180				Marker Pr234				Marker Pr297	
	bp64	bp155	bp159	bp194	bp109	bp164	bp200	bp205	bp174	bp239
A (NT)	C	C	A	A	G	C	T	T	G	A
46_B08 (NT)	C	C	A	A	G	C	T	T	G	A
D (T)	C/T	C/G	A/G	A/C	G/T	C/T	A/T	G/T	G/A	Het
46_B06 (T)	C/T	C/G	A/G	A/C	G/T	C/T	A/T	G/T	G/A	Het
25_E11 (T)	C	C	A	A	G	C	T	T	G	A
25_E01 (NT)	C	C	A	A	G	C	T	T	G/A	Het

The recombination breakpoints in the recombinants genotyped with the markers are shown in table 5.6 in terms of whether the genotype at each marker was associated with copper tolerance or non-tolerance in the parent plants. Looking at these breakpoints, the key informative individuals here are the NTR 25_E01, which has marker genotypes associated with tolerance until the Pr234 marker meaning recombination occurs between Pr297 and Pr234, and the NTR 19_H12 which has the same genotype as 25_E01 yet crosses between this plant and Cop do not cause necrosis. So, the association of the necrosis phenotype with tolerance may or may not be disrupted during the recombination event occurring in between the Pr297 and Pr234 loci. As such, the necrosis locus must be positioned in between these two markers as if it lay on the other side of either marker then the relationship between tolerance and necrosis would either always be disrupted (if *NEC* was between G5.2nd and Pr297) or never disrupted (if *NEC* was between Pr234 and Pr180).

As such, these two SNP markers, Pr297 and Pr234, demarcate two new limits of the recombination window containing the *NEC* locus and have narrowed the window down to a 59kb region.

5.3.2 Bioinformatics

Due to the robust evidence supporting a role for classical *R* genes in hybrid necrosis, the first step in the initial bioinformatics analysis was to establish the presence or absence of any such genes in the window defined by SNP markers Pr297 and Pr234. The entire 1.1Mb of scaffold_84 with the *R* genes and *T* locus marker positions mapped is shown in figure 5.6 and illustrates that the nearest *R* gene position is at 630kb making it some 330kb outside the window containing *NEC*. These data therefore establish that there are no known *R* genes in the window containing the *NEC* locus and consequently identifying other candidates for *NEC* required identification and assessment of all the other predicted genes/mRNAs within this region.

Table 5.6 The marker genotypes in parent plants A and D and in TRs and NTRs with regards to their associated tolerant/non-tolerant phenotype thereby revealing the recombination breakpoints in the D x A plants recombinant between e242 and T. The three TRs shown in grey print have genotypes at Atg5 that are inconsistent with the other data (as described in Chapter 4). The informative SNP genotypes in terms of narrowing the region containing NEC are shown in bold italics.

[T = copper tolerant; NT = non-tolerant; L = lethal/necrotic; NL=non-lethal/necrotic]

Plant	Phenotype for:		Marker Genotype								
	Copper tolerance	Hybrid Necrosis	e242	Sg1788	e481	G5.2nd	Pr297 (DUF)	Pr234 (JmjC)	Pr180	G4.l1	Atg5
D	T	L	T	T	T	T	T	T	T	T	T
TRs											
25_A09	T	L	NT	NT	NT	-	NT	NT	NT	-	NT
43_H01	T	L	NT	NT	NT	-	NT	NT	NT	-	NT
28_G06	T	L	NT	NT	NT	-	NT	NT	NT	-	NT
25_E11	T	NL	NT	NT	NT	NT	NT	NT	NT	T	T
26_C07	T	L	NT	NT	NT	-	NT	T	T	-	T
26_H02	T	L	NT	NT	NT	-	NT	T	T	-	T
32_G02	T	L	NT	NT	NT	?	T	T	T	-	T
34_A03	T	L	NT	NT	NT	-	T	T	T	-	T
35_F09	T	L	NT	NT	NT	-	?	T	T	-	T
A	NT	NL	NT	NT	NT	NT	NT	NT	NT	NT	NT
NTRs											
19_C03	NT	NL	T	NT	NT	-	-	-	-	-	NT
19_F06	NT	NL	T	NT	NT	-	-	-	-	-	NT
35_A12	NT	NL	T	NT	NT	-	-	-	-	-	NT
24_D09	NT	NL	T	T	T	T	NT	NT	NT	-	NT
35_B06	NT	NL	T	T	T	-	NT	NT	NT	-	NT
48_B01	NT	NL	T	T	T	-	NT	NT	NT	-	NT
19_H12	NT	NL	T	T	T	-	***T***	***NT***	NT	-	NT
25_E01	NT	L	T	T	T	T	***T***	***NT***	NT	NT	NT
32_H09	NT	NL	T	T	T	-	?	NT	NT	-	NT
10_F08	NT	NL	T	?	NT	-	?	NT	NT	-	NT

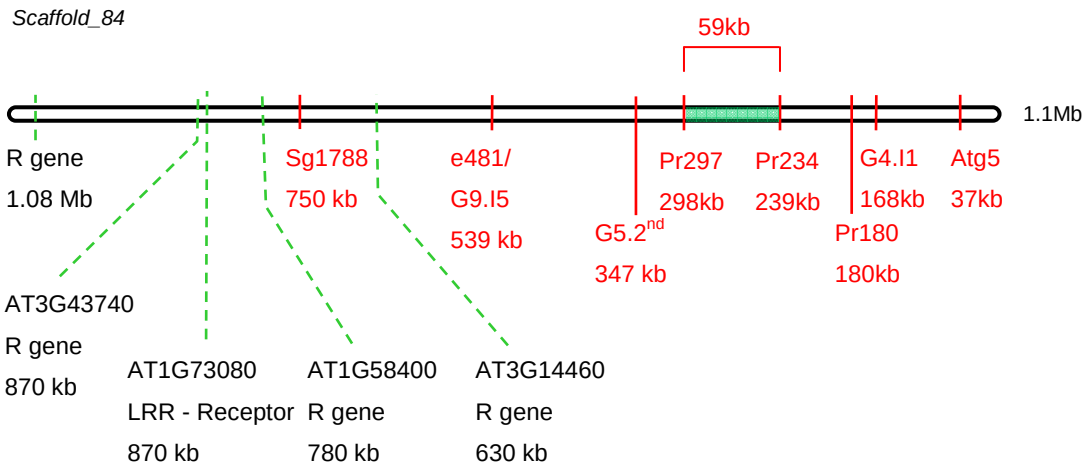


Figure 5.4 The position on scaffold_84 of the Mimulus genome of known *R* genes, the *T* locus markers identified in Chapters 4 & 5 and the recombination window containing *NEC*.

The 59kb window contains three mRNA transcripts, as shown in figure 5.5, two of which were the genes on which the SNP markers Pr297 and Pr234 were based. At ~298kb is a gene transcript with putative conserved domain belonging to the DUF563 glycosyltransferase protein superfamily. This is a family of uncharacterised proteins (Pfam 04577) whose molecular function (GO:0016757) is the transferring of glycosyl groups from one compound to another during different metabolic processes, and as such are involved in metabolic processes. At ~239kb lies a gene encoding an mRNA with a putative conserved domain belonging to the Jumonji (JmjC) protein superfamily. Both the Pfam (PF02373) and NCBI BlastP description of this domain state that “The JmjC domain belongs to the Cupin superfamily. JmjC-domain proteins may be protein hydroxylases that catalyse a novel histone modification”. At ~282kb is an mRNA encoding a transcript that is not annotated and therefore is of unknown function.

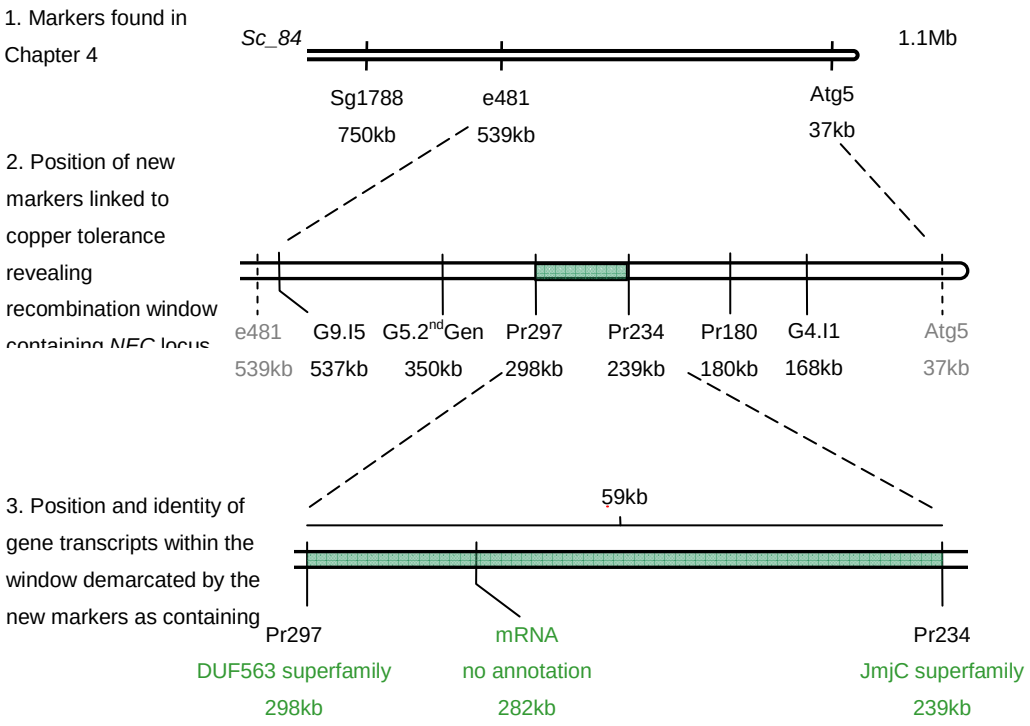


Figure 5.5 A summary of the results of the mapping, fine mapping and bioinformatics in Chapters 4 & 5 showing the informative markers on scaffold_84 of the Mimulus genome that narrow down the region containing NEC and the three candidate genes within this 59kb window.

5.4 Discussion

The genomic location of the *NEC* gene of Copperopolis which is incompatible with genes from Cerig when brought together in the genome of a hybrid has been identified on chromosome 9, lying within a 59kb window that contains only three genes. A combination of bioinformatics and molecular genotyping of recombinants identified five markers that enabled the refinement of the region containing the *NEC* gene from the previously identified 500kb window to this 59kb window containing the three genes: a DUF563 glycosyltransferase, a Jumonji-domain protein and a gene with no functional annotation.

Previous studies on hybrid necrosis implicated the role of plant disease resistance genes (*R* genes) in this form of hybrid incompatibility and as the phenotype of the hybrids in this project was very similar to that of the *R* gene hybrid incompatibilities in the literature, it was considered likely that the *NEC* gene in Copperopolis could be an *R* gene. Bioinformatic analysis on the

region containing the Cop *NEC* gene has revealed that there are no *R* genes in the region and the two annotated genes have not been described in other hybrid necrosis systems. Thus, either (i) a gene with a role in the immune response other than an *R* gene is involved or (ii) a gene with a different biological function than in disease resistance is capable of producing the necrotic phenotype in hybrids or (iii) the complementary genes from Cerig with which *NEC* interacts are involved in the disease response. Attempting to determine how likely the genes within the region containing *NEC* are to be the incompatiblity gene causing necrosis depends on understanding more about their biological functions.

The Jumonji-domain protein at 239kb

The Jumonji (Jmj)-domain protein family was first described as a group of transcription moderators possessing several different kinds of DNA-binding domains, including the highly conserved JmjC domain at the C-terminus (Balciunas and Ronne, 2000). Since then, evidence has accumulated to prove the role of JmjC-domain-containing proteins in gene expression regulation through their action as histone demethylases (Klose et al., 2006).

In eukaryotic cells, DNA is packaged as chromatin which consists of a series of nucleosomes in which 147 base pairs of DNA is wrapped around an octamer of four core histones (H2A, H2B, H3 and H4) in either a heterochromatic (condensed or “closed”) or euchromatic (extended or “open”) state. Generally, gene transcription is repressed when DNA is in the heterochromatin state and increased in the euchromatin state as this is more accessible to transcription factors. However, histone modification through ubiquitination, methylation, acetylation and phosphorylation can activate or repress transcription by generating a more open or closed chromatin configuration. The effect on gene expression of the modification depends on which histone amino acid is modified, the type of modification and the degree of modification (e.g. monomethylation, dimethylation, trimethylation) (Pfluger and Wagner, 2007). In this way, histone modification has a major role in regulating gene expression and so, indirectly, do the enzymes that perform the modifications.

Most of the literature on JmjC-domain containing proteins relates to their function as histone demethylases. Histone methylation, the addition of one, two or three methyl groups to an amino acid within the histone protein, is an important form of gene regulation and is regulated by enzymes that add or remove methyl groups. The JmjC-domain-containing histone demethylases (JHDMs) catalyses lysine demethylation (Klose et al., 2006) and can remove all three histone lysine-methylation states. The effect of lysine methylation can be either to activate or repress transcription depending on which lysine molecule is involved: methylation of H3

lysine (K) 4 or K36 causes activation whilst methylation of H3 K9, K27 or H4 K20 causes repression (Pfluger and Wagner, 2007). As such, the demethylation activity of JHDMs could also accordingly cause gene expression activation or repression. In *Arabidopsis*, H3K9 and H4K20 are both methylated and closed whilst H3K36 is methylated and open and JmjC-domain histone demethylases have been found causing both activation and repression by demethylating these histones (Pfluger and Wagner, 2007). Paramount to predicting the effects of a JHDM is to know which lysine residue it acts upon and whether it results in gene expression or repression.

Previous studies have revealed a methyltransferase domain-containing protein that is involved in hybrid male sterility in crosses between two subspecies of mice due to a DM genic incompatibility (Mihola et al., 2009). In their paper on *Prdm9*, Mihola et al (2009) found this gene encoded a methyltransferase that epigenetically modifies a chromatin structural protein, histone H3, which activates the promoters of genes essential for meiosis. Although as of yet no complementary gene has been found from the second subspecies of mouse, it would seem that in hybrids the normal functional form of the gene cannot work and so meiotic defects occurred including abnormalities in gene expression, chromosomal pairing, and male meiotic sex chromosome inactivation. However, other evidence now implicates not the methyltransferase domain but the zinc finger domains as the source of incompatibility that manifests itself as sterility in the hybrid. This suggests the essential role of Prdm9 is not its methyltransferase activity but the affinity of the zinc finger domain in binding to rapidly-evolving satellite DNA sequences and suppressing their meiotic drive and associated deleterious effects (Oliver et al., 2009).

Given the role of JmjC-domain proteins as histone demethylases in gene regulation, this could be a strong candidate for the *NEC* gene. However, the function of this gene in *Mimulus* is not known and it could be that it is a transcription modifier with another function other than histone modification which could impact on its strength as a candidate. All things considered, this seems an attractive *NEC* candidate gene in theory but with little supporting empirical evidence.

The DUF563 Glycosyltransferase at 298kb

Glycosyltransferases are encoded by large multigene families sometimes comprising several hundred genes. Most glycosyltransferases are very specific and, given that there are estimated to be more than one-hundred distinct glycosidic linkages in a cell, it is likely that each different linkage requires the action of a distinct glycosyltransferase. Indeed, in plants there are sixty-nine families of glycosyltransferase that are classified on the basis of substrate recognition and sequence relatedness (Lim and Bowles, 2004). This specificity is also reflected in the

Arabidopsis genome which possesses hundreds of putative glycosyltransferase genes, although the biochemical activity of the products of these genes is largely unknown (Keegstra and Raikhel, 2001).

Glycosyltransferases are involved in a variety of processes including the biosynthesis of cell-wall polysaccharides and glycoproteins as well as the modification of secondary metabolites such as hormones and flavonoids (Gachon et al., 2005, Lim et al., 2002, Lim and Bowles, 2004). Perhaps more interesting, however, is that glycosyltransferases perform at least two important roles in the plant disease response.

First, the glycosylation of pattern recognition receptors that activate a first line of defence against pathogens can be essential for their stability and function (Haweker et al., 2010). For example, the tomato disease resistance protein Cf-9, which has been identified as causing hybrid necrosis (Wulff et al., 2004), has many glycosylation sites that contribute to its activity, with those in the leucine-rich repeat domain being essential, and mutation of these sites compromises the ability of Cf-9 to trigger Avr 9-mediated PCD (van der Hoorn et al., 2005). In *Cf* gene products, amino acid differences alter potential glycosylation sites (Michelmore and Meyers, 1998) making it possible that divergence in a Cerig individual at the target glycosylation site for this glycosyltransferase could prevent this protein functioning in the hybrid genome. In addition, for several mammalian extracellular receptors, the pattern of *N*-glycosylation influences ligand binding (Michelmore and Meyers, 1998). If a similar affect occurs in plants, it could cause *R* genes to no longer recognise pathogen-derived molecules. However, this glycosylation activity seems improbable as the causative factor of hybrid necrosis as anomalous glycosylation of disease response gene products appears to reduce their ability to activate the defence response. Then again, it may be possible that deviant glycosylation patterns could cause *R* gene activation by aberrant ligand binding as in a mechanism model proposed by Yamamoto et al (2010) whereby hybrid necrosis is caused when the casein kinase encoded by *hbd2* acts like the Avr protein, or a protein disturbed by the Avr protein, which is recognised by one or more *R* genes in *hbd3* and triggers the immune response.

Another way in which aberrant glycosylation could cause activation of the disease response is via the guard hypothesis (van der Biezen and Jones, 1998). This model proposes that pathogen effector proteins and R proteins do not interact directly. Instead, R proteins are associated with guardee proteins, host proteins that are the target of a pathogen effector protein, and are activated by interactions between the pathogen Avr protein and the guardee protein. For example, in *Arabidopsis*, the guardee protein RIN4 prevents the inappropriate activation of the R proteins RPM1 and RPS2. However, when a pathogen effector interacts with RIN4, RPM1 and RPS2 are activated and elicit the disease resistance response. What is particularly

interesting is that the R proteins are activated by hyperphosphorylation of RIN4 upon effector interaction (Belkhadir et al., 2004). It may be possible that glycosylation patterns on such guardee proteins could also be important for negatively regulating R protein activity, and so the incorrect glycosylation causes R protein activation in the absence of a pathogen. Interestingly, the aforementioned Cf-9 R protein, identified as being involved in hybrid necrosis (Wulff et al., 2004) is thought to function via the guard hypothesis (Dangl and Jones, 2001).

A second role for glycosyltransferases in the immune response is through the regulation of salicylic acid (SA) homeostasis (Gachon et al., 2005). A key role of SA is as the endogenous signaling molecule that mediates the plant defence response by activating disease resistance that is frequently associated with the hypersensitive response (HR) characterised by localised programmed cell death (PCD) (Dangl et al., 1996, Klessig and Malamy, 1994, Raskin, 1992). SA appears to control the timing and extent of cell death and enhances the defence responses by promoting lesion formation and inducing systemic acquired resistance (SAR) (Alvarez, 2000). Although the significance and implications of plant hormone modifications remain speculative, in general, conjugation with a glycosyl moiety is considered an inactivation mechanism (Gachon et al., 2005) and it is thought that glucosylation represents detoxification of SA (Lim et al., 2002). As such, if this glycosyltransferase has a specific function to glucosylate SA and thereby inactivates the SA signal that normally elicits the HR, the inability of this enzyme to perform its function in the genome of a hybrid could prevent the SA signal being inactivated and therefore the necrosis phenotype would result from the SA-induction of the HR and programmed cell death. If glycosylation of SA is important in inactivating its signalling function, then the specificity of the glycosyltransferase could be of vital importance if it cannot function. However, the very specificity also lessens the probability that this gene, out of possibly hundreds, encodes a glycosyltransferase that functions alone to inactivate SA.

Additional evidence for the importance of the SA pathway in driving hybrid incompatibility is demonstrated by Alcazar et al (2009) who found that the severe growth defects seen in hybrids depended on activation of the SA stress signalling pathway and that levels of SA were dramatically higher in hybrids where the incompatibility phenotype was most apparent. One of the interacting loci was mapped to a TIR-NB-LRR RPP1-like cluster, homologs of which are known to recognize specific pathogen effectors and trigger SA-dependent defences. The authors conclude that multiple loci with small effects, or other complex epistatic networks not yet identified, act as modifiers of the incompatible phenotypes and that different epistatic networks appear capable of promoting hybrid incompatibilities through environmentally conditioned activation of plant immune responses. Not only does this provide additional evidence for the involvement of the plant immune system in causing hybrid incompatibilities but it also supports a

potential role for other genes of the plant immune system, and not solely *R* genes, in hybrid necrosis.

Another important role of glycosylation is in the correct folding of proteins in the endoplasmic reticulum. Prevention of *N*-glycosylation induces the unfolded protein response which, interestingly, initiates PCD (Urade, 2007). If this DUF563 glycosyltransferase is responsible for the specific glycosylation of a conserved site in a ubiquitous protein, or several proteins given that there is more than one gene involved in Cerig, then its inability to function in the hybrid genome could potentially result in widespread PCD.

Clearly, glycosyltransferases perform important biological functions in a wide variety of processes so the likelihood of this gene being involved in a DM incompatibility resulting in hybrid necrosis must surely depend on the biochemical activity of the protein. At present the specific activity of this glycosyltransferase is unknown making it extremely difficult to accurately assess its potential involvement in incompatibility. Nevertheless, there is compelling information, mainly from disease resistance, that this could be a strong candidate for *NEC*.

The mRNA with no functional annotation at 282kb

Assessing the potential for this gene as *NEC* is difficult as this gene is not annotated on the *Mimulus* genome sequence at www.phytozome.net and an NCBI blast on both the mRNA and peptide sequence indicates there are no putative conserved domains. However, using an open reading frame (ORF) finder (http://www.ncbi.nlm.nih.gov/projects/gorf/) one ORF was found that contains a non-specific conserved domain belonging to the exonuclease/ endonuclease/ phosphatase superfamily often involved in intracellular signalling.

There is evidence that protein phosphatases play a role in defence signal transduction (Stone, 1994 cited by Bent, 1996). Indeed, as described for the glycosyltransferase gene, the activation of *Arabidopsis* R proteins in response to *P. syringe* relies on the hyperphosphorylation of the RIN4 guardee protein (Belkhadir et al., 2004). However, the literature on the role of phosphorylation in plant immune responses usually concerns the role of protein kinases. Protein kinases are known to regulate the majority of cellular pathways, especially those involved in signal transduction. Members of the serine/threonine protein kinases group, the mitogen-activated protein kinases (MAPK), respond to extracellular stimuli and regulate various cellular activities, such as gene expression, mitosis, differentiation, proliferation and, more pertinently, apoptosis.

There is already evidence in the literature that protein kinases can act as DM genes. In the platyfish *Xiphophorus maculates*, the DM gene *Xmrk2* encodes a tyrosine kinase and hybrids between *X. maculates* and *X. helleri*, the green swordtail, develop lethal tumours. Furthermore, there is indication for the involvement of kinases in hybrid lethality, and particularly necrosis, based on the knowledge that the activation of members of the serine/threonine protein kinases group, the mitogen-activated protein kinases (MAPKs), is one of a plants' earliest responses to pathogens and induces the expression of defence genes and elicits the HR (Zhang and Liu, 2001). For example, studies in tobacco have demonstrated that the activation of the MAPK salicylic acid-induced protein kinase (SIPK) by an additional upstream MAPK kinase (MAPKK) leads to HR-like cell death (Hailing et al., 2003). Once again, this concurs with the similarities seen between the plant immune responses of HR/programmed cell death and hybrid necrosis, furthering the theory that hybrid necrosis is elicited by aberrant interactions of genes involved in plant defence.

Given the role protein phosphatases have in defence signal transduction, the requirement for hyperphosphorylation to activate R proteins, and the known involvement of a kinase in hybrid lethality, this gene could also be considered a good candidate for *NEC*, although its unapparent putative conserved domain and lack of functional annotation in the *Mimulus* genome sequence should not be disregarded.

Several attempts were made to amplify up the entire sequence of each of these three mRNAs with the intention of sequencing so as to identify variation depending on whether the parent was copper tolerant (Cop and D) or non-tolerant (Cerig and A). However, despite repeated attempts this proved to be unachievable due to difficulties incurred in the PCR amplification of the genes, primarily due to problems with primer design due to sequence similarity between the desired fragment and sequences on other scaffolds. Despite repeated PCRs, primer redesign, amplifying the whole mRNA sequence or sections of the sequence, the sequence similarity proved such a problem that ultimately this work was abandoned. However, as exons from all three of the mRNAs were used to successfully develop SNP markers, there is clearly variation in the sequences of all these genes and so sequencing might not have revealed any conclusive data, based on sequence variation, as to which was most likely to be the *NEC* gene.

5.5 Conclusion

Although a number of loci have been identified as causing hybrid necrosis, relatively few have been identified in terms of their biological function. Currently, those with known function are all implicated as being involved in the disease response (Alcazar et al., 2009, Bomblies et al., 2007, Herbers et al., 1996, Jeuken et al., 2009, Jiang W et al., 2008, Kruger et al., 2002, Lamb and Dixon, 1997, Wulff et al., 2004, Yamamoto et al., 2010). Consequently, identifying more hybrid necrosis genes will provide valuable insights into the types of genes involved in causing hybrid dysfunctions and may either consolidate the evidence for the role of the plant immune system in hybrid necrosis or provide an example of the involvement of a gene with an entirely different function.

The locus in the Copperopolis population of *M. guttatus* that interacts with a small number of genes in the Cerig population to cause hybrid necrosis has now been mapped to a 59kb region on Chromosome 9 containing three candidate genes. Without specific information about the expression or function of these three genes, assessing which is most likely to be *NEC* is speculative, especially as all three have evidence to support their potential role in an incompatibility interaction. Interestingly, although none of these genes are actual disease resistance genes, at least two have potential roles in the disease response indicating that this crossing barrier between Cop and Cerig could be the result of aberrant immune activation. Further fine mapping of the 59kb window could ultimately result in only a single candidate gene in the window containing *NEC* thereby enabling identification of the locus and its function. The future work discussed in Chapter 7 provides some methods that could be pursued concerning the *NEC* gene.

Chapter 6

QTL mapping of the genes in the Cerig population that interact with the NEC gene in Copperopolis to cause hybrid necrosis

6.1 Introduction

Hybrid necrosis in the F_1 offspring from crosses between copper-tolerant 'Copperopolis' (Cop) and copper non-tolerant 'Cerig' populations of *Mimulus guttatus* varies in the proportion of hybrids affected depending on the individual Cerig parent involved. In the initial investigation into this crossing barrier Macnair and Christie (1983) found that in some crosses the proportion of offspring affected by hybrid necrosis was almost one-hundred per cent whilst in other crosses it was almost zero. This wide distribution in the degree of the phenotype was confirmed by the characterisation work in Chapter 3. As the hybrid necrosis is brought about by a Dobzhansky-Muller (DM) genic incompatibility and there is known to be a single locus, *NEC*, involved from the Cop population (Macnair and Christie, 1983; Chapters 4 & 5 of this thesis), the variation in the degree of necrosis signifies that the Cerig population is polymorphic for the genes involved so that the genotypes *aa*, *Aa* and *AA* all exist within the population for one or more of these genes (Macnair and Christie, 1983) with incompatibility occurring between the *A* allele from Cerig and the *NEC* locus from Cop. Furthermore, Macnair & Christie (1983) found that, based on the number of classes of progeny (i.e. the number of classes of necrosis), there must be at least two and probably more genes controlling the Cerig side of the barrier but the spread of ratios also suggested that there cannot be very many. This conclusion is supported by the results reported in Chapter 3 of this thesis which also indicate the involvement of a small number of genes, which, given the number of classes of necrosis caused by the different Cerig parents, are likely to be fewer than five. Allelic polymorphisms at genes involved in DM incompatibilities resulting in variation in the number of necrotic hybrids have been found in other systems including *Crepis* (Hollingshead, 1930), *Gossypium* (Stelly, 1990) and *Nicotiana* (Bomblies and Weigel, 2007).

In addition to variation in the degree of the phenotype, the severity of the hybrid necrosis symptoms in crosses between Cop and Cerig ranges from weak (yellowing) to severe (necrotic, severely dwarfed, lethal) (see Chapter 3 of this thesis). From other studies on plant hybrid incompatibilities, it is evident that this wide variation in the severity of the incompatibility phenotype is not uncommon, and is typically attributed to a gene dosage effect in which the incompatibility phenotype increases in proportion to the number of incompatible alleles present. Such gene dosage effects have been reported in hybrid incompatibilities in *Gossypium* (Song et

al., 2009), rice (Ichitani et al., 2007), *Arabidopsis* (Bomblies et al., 2007), and wheat (Hermsen, 1963, Chu et al., 2006). This gene dosage effect might be more accurately referred to as an allele dosage effect as it is the number of dominant incompatible alleles that affect the severity of the incompatibility phenotype. A number of additional factors are also reported to affect the severity of the necrosis symptoms including allelic expressivity, where the DM-genes have multiple alleles each of which differ in the degree of sterility or lethality they bring about (Hermsen, 1963), and the time of onset of the first symptoms where the earlier the first symptoms appear, the more severe the phenotype is (Hermsen, 1963, Macnair and Christie, 1983, Song et al., 2009).

Although examples of hybrid incompatibilities and the genes controlling these crossing barriers are increasing in number, most comprise two-locus interactions and still relatively little is known about interactions involving three or more loci. In these more complex interactions, it might be necessary for there to be an incompatible allele at one locus before an incompatible allele at a second locus has an effect on the phenotype or there may be a requirement for particular alleles at three or more loci before fertility or viability is affected (Coyne and Orr, 1998, Coyne and Orr, 2004, Orr, 1995). For example, in female hybrid sterility from crosses between two rice cultivars there is a requirement for homozygosity for the incompatible allele at one locus and at least one incompatible allele at both of the other two loci before sterility occurs (Kubo and Yoshimura, 2005). Furthermore, this three-locus incompatibility shows an allele dosage effect as females can be fully sterile, semi-sterile or fully fertile depending on the number of incompatible alleles they receive (Kubo and Yoshimura, 2005).

The nature of how the incompatibility interaction between the single locus in Cop (*NEC*) and the small number of genes in Cerig brings about hybrid necrosis is unknown. It could require one incompatible allele at each of the DM loci in Cerig for necrosis to occur or there may be an additive/allele dosage effect so that a single incompatible allele at one locus is sufficient to cause some necrosis and the degree of necrosis increases with each additional locus that possesses an incompatible allele. As there are a small number of polymorphic genes involved in the Cerig side of the crossing barrier (Macnair and Christie, 1983) the strength of the barrier is likely to depend on the specific arrangement of alleles at all these loci.

The majority of research into the genetics of the incompatibility between Cop and Cerig has, up to now, predominantly focused on the single incompatibility locus, *NEC*, in the Cop population (Christie and Macnair, 1984, Christie and Macnair, 1987, Macnair and Christie, 1983). Indeed, it appears that it is not unusual for studies on postzygotic reproductive isolating barriers resulting from DM genic incompatibilities to focus on the gene(s) from just one side of the barrier as, despite the growing number of DM genes being identified (see Chapter 1, table 1.4), relatively few studies identify the interacting genes from both of the populations/species involved. Those

that have been described are summarised in table 6.1. As such, a preliminary investigation into identifying the loci of the genes in Cerig that interact with the *NEC* gene in Cop can begin the process of elucidating the unknown side of this partial RI barrier and thereby offer a valuable additional example of interacting DM genes.

Furthermore, most of the examples in which interacting DM loci have been identified have involved the interaction of two incompatible genes, probably due to the relatively simplistic nature of these interactions and the comparative ease of identifying only two genes, although even these relatively simple systems can be complicated through the activity of multiple alleles (Hermsen, 1963) or the requirement for three specific alleles at the two loci (Lee, 1981). In plants, there currently seems to be only one example of an incompatibility involving more than two loci: that of the aforementioned female rice sterility brought about by the interaction of incompatible alleles at three loci (Kubo and Yoshimura, 2005). As the incompatibility between Cop and Cerig involves a small number of genes from the Cerig population (Macnair & Christie, 1983), it provides a potential example of a multi-gene incompatibility interaction and so identifying the genes in Cerig would be a valuable contribution to the understanding of the more complex interactions between incompatible loci.

The aim of this section of research is to begin the process of identifying the DM genes in Cerig by locating quantitative trait loci that are associated with the hybrid necrosis phenotype. Two strategies for achieving this were considered. One was the traditional approach of screening for random markers distributed throughout the genome. The second was based on targeting markers for plant disease resistance genes (*R* genes).

There is an accumulating number of studies revealing the involvement in hybrid necrosis of the plant immune system acting in an autoimmune-type response brought about by hybrid incompatibilities involving *R* genes (Bomblies, 2009, Bomblies et al., 2007, Bomblies and Weigel, 2007, Jeuken et al., 2009, Kruger et al., 2002, Masuda, 2007, Wulff et al., 2004, Yamamoto et al., 2010). Although genes with functions other than as *R* genes have been identified as causing hybrid necrosis, even these are involved in an autoimmune response as they interact with *R* genes thereby bringing about necrosis. An example of this is found in rice where *hbd2*, a DM-gene in *indica* rice encoding a casein kinase (Yamamoto et al., 2010) is incompatible with an *R* gene, *hbd3*, from *japonica* rice (Matsubara et al., 2007, Yamamoto et al., 2007, 2010) meaning hybrid necrosis in this cross is via an autoimmune-type response. A similar mechanism could be in place in *M. guttatus*: the incompatible (*NEC*) gene in Cop was found not to be an *R* gene (see chapter 5) but it is possible that the *NEC* allele could interact with alleles of *R* genes from Cerig. Given the evidence for the role of *R* genes in hybrid necrosis and the strong phenotypic similarity of hybrid necrosis in other studies and in this project, and the fact that *NEC* is not an *R gene*, it was decided that a more targeted screening for markers

linked to *R gene*s could prove to be an effective approach as it would focus on regions of the genome with high potential involvement in the phenotype.

Table 6.1 Identified pairs or sets of complementary genes whose diverged alleles interact in a Dobzhansky-Muller-type incompatibility resulting in hybrid dysfunction.

Species 1		Species 2			
Species	**Gene**	**Species**	**Gene**	**Hybrid…**	**References**
D.simulans	*Lhr*	*D.melanogaster*	*Hmr*	Larval lethality	(Brideau et al., 2006, Hutter and Ashburner, 1987, Watanabe, 1979)
D.simulans	*Mhr*	*D.melanogaster*	*Zhr*	Embryonic lethality	(Ferree and Barbash, 2009, Sawamura, 2000, Sawamura et al., 1993a, 1993c)
G.hirsutum	Le_3	*G.barbadense* R4-4	Le_4	Lethality	(Song et al., 2009)
Triticum spp	*Ne1*	*Triticum* spp	*Ne2*	Necrosis	(Chu et al., 2006, Hermsen, 1963, Singh et al., 2000)
L. pimpinellifolium	*Cf-2*	*L.esculentum*	*RCR3*	Necrosis	(Kruger et al., 2002)
A.thaliana Uk-1	*DM2*	*A.thaliana* Uk-3	*DM1*	Necrosis	(Bomblies et al., 2007)
O. sativa cv Jamaica	*Hwc1*	*O. sativa* cv Norin 8, Taichung 65, Nipponebare	*Hwc2*	Necrosis	(Ichitani et al., 2001, 2007, Saito et al., 2007, Sato and Morishima, 1987)
O. sativa L. ssp *indica* Habataki cultivar	*hbd2*	*Oryza sativa* spp. j*aponica* Koshihikari & Sasanishiki cultivars	*hbd3*	Necrosis (breakdown)	(Matsubara et al., 2007, 2007, Yamamoto et al., 2010)
O. sativa India/Japanese varieties	*hwb1*	*O. sativa* India/Japanese Varieties	*hwb2*	Necrosis (breakdown)	(Oka, 1956)
O. sativa Thai cultivar Col. No 15	*hwd1*	*O. sativa* Nepalese cultivar Siborunauli 1	*hwd2*	Necrosis (breakdown)	(Fukuoka et al., 1998)
O. sativa L. ssp *indica* IR24	*hwe1*	*O. sativa* L. ssp *japonica* variety Asominori	*hwe2*	Necrosis (breakdown)	(Kubo and Yoshimura, 2002)
O. sativa L. ssp *indica* Milyang 23	*hwh1*	*O. sativa* L. ssp *japonica* Tong 88-7	*hwh2*	Necrosis (breakdown)	(Jiang W et al., 2008)
O. sativa L. ssp *japonica* variety Asominori	*hsa2-As & hsa3-As*	*O. sativa* L. ssp *indica* variety IR24	*hsa1-IR*	Female sterility	(Kubo and Yoshimura, 2005)
Oryza sativa spp. *indica*	SaM^+SaF^+	*O.staiva* L. ssp *japonica*	SaM^-	Male sterility	(Long et al., 2008)
S. bayanus	*AEP2*	*S. cerevisiae*	*OLI1*	Sterility	(Lee et al., 2008)

6.2 Methods and Materials

A more detailed description of the techniques and plants used in this section is given in Chapter 2.

6.2.1 QTL analysis and Bulked Segregant Analysis

Quantitative trait loci (QTL) analysis is usually associated with a segregating population of plants in which each individual has to be phenotyped for the trait of interest and genotyped with all the selected molecular markers. A molecular marker that shows polymorphism between the parents of the population and is closely-linked to a QTL regulating a particular trait will mainly co-segregate with that QTL.

Genotyping each individual in a segregating population of sufficient size (two-hundred individuals (Quarrie et al., 1999) is both laborious and costly in terms of consumables. For example, to genotype this population with, say, one-hundred markers would require 20,000 PCR reactions to be performed and analysed. However, a modified version of QTL analysis termed bulked segregant analysis (BSA), first described for use in plant genetics by Michelmore et al (1991), permits the rapid identification of markers linked to any specific gene or genomic region without the need to genotype every individual.

In this BSA method, plants from a segregating population are grouped according to their phenotypic expression of a trait and the DNA from plants at the two phenotypic extremes is pooled, resulting in only two DNA samples to genotype for the markers enabling those one-hundred markers to be screened in just 200 PCR reactions. Markers that are polymorphic between the pools will be genetically linked to the loci determining the trait used to construct the pools (Michelmore et al., 1991). Consequently, if the chromosomal locations of the molecular markers are known, the map location of closely-linked QTLs can be deduced without having to genotype every individual in a segregating population (Michelmore et al., 1991, Quarrie et al., 1999). As such, BSA allows the rapid screening of many loci and therefore the identification of segregating markers in the target region (Michelmore et al., 1991) and so makes it possible to screen more markers. Those markers that are polymorphic can then be tested further in the other individuals that are not at the extreme phenotypes of the segregating population.

For the research undertaken in this chapter, BSA is a suitable method as this was a first-pass analysis for the QTLs for hybrid necrosis and the expectation was for a small number of genes of large effect (Macnair & Christie, 1983). Using this method had the added advantage of making it possible to screen and analyse a larger number of markers in a short period of time.

6.2.2 Generating a segregating F_2 population for BSA

Performing QTL analysis required an F_2 Cerig population segregating for the hybrid necrosis trait. This population was generated by first performing crosses between the two Cerig plants previously identified as causing the highest and lowest rates of hybrid necrosis when used in crosses with Cop (see Chapter 3). These were Cer10 and Cer35 respectively. A single F_1 progeny from this cross was selfed and seed from this cross were grown to provide the segregating F_2 population. Individuals from the segregating F_2 population were then tested for their phenotypic expression of the hybrid necrosis trait by performing crosses to Cop and scoring the progeny for the number of necrotic hybrids.

A modified version of the BSA method was used. Instead of pooling the DNA from the F_2s at the two extreme phenotypes, the DNA from the twenty-four individuals causing the highest and lowest hybrid necrosis was tested individually with markers. The reason for this modification was that the hybrid necrosis rate varied from around 40% to 75% at the high-necrosis extreme, presumably due to genetic variation, and so pooling the DNA in the usual BSA method could distort the results and prevent the detection of linked markers. Having twenty-four genotypes for each marker, instead of the two genotypes obtained through pooling DNA in a true BSA approach, would make it more evident if a marker was co-segregating with hybrid necrosis and also reduce the effect that any errors in the scoring of the hybrid necrosis phenotype might have on marker identification.

6.2.3 Phenotyping the segregating F_2 population for hybrid necrosis

A total of two-hundred plants were grown for the segregating F_2 population generated in 6.2.2. Establishing the degree (proportion) of hybrid necrosis of each F_2, hereafter referred to as their hybrid necrosis rate, required crossing each individual with Cop so as to generate the hybrid progeny which were grown and scored for the number that displayed the hybrid necrosis phenotype. Of the 200 F_2 grown, 168 were successfully used in crosses with Cop. The remaining thirty-two F_2s could not be used due to mortality prior to successful crosses, failure to produce inflorescences or poor seed production, all primarily caused by pest-associated problems.

Two levels of replication were performed when establishing the hybrid necrosis rate caused by each F_2. One level of replication was within each F_2 x Cop cross at the “tray-level” and was performed in order to take into account a possible effect of the position of a tray on the benches. For each cross, a total of sixty seedlings were transferred into two trays each containing thirty plants so that were two replicate trays for each cross. Using the RAND function of Microsoft

Excel, the trays were assigned a random position on one of three adjacent benches with operational blackouts so that day length was restricted to nine hours. A total of twelve crosses had only one tray (i.e. no tray-level replication) due to low seed production or germination, the cause of which was uncertain.

The second level of replication was accomplished by sowing two different crosses for each F_2 and so is termed the "cross-level". Of the 168 F_2s successfully crossed with Cop, five had three replicate crosses, 148 had two replicate crosses and fifteen were without a replicate cross due to poor seed production or low germination. This level of replication was performed so as to reduce the occurrence of errors in the hybrid necrosis rates by verifying the rate of each F_2 in independent crosses. In addition, where possible the cross-level replicates were not sown at the same time in order to take into account any effects of the growing environment and so was an improvement on the method used by Macnair & Christie (1983) which did not take into account environmental effects. The growing environment, particularly the ambient temperature, is an important factor since hybrid necrosis has been shown to display a temperature sensitive response with the phenotype being alleviated at higher temperatures (Alcazar et al., 2009, Bomblies et al., 2007, Jeuken et al., 2009, Masuda et al., 2007, Saito et al., 2007). This form of replication also lessened the potential impact of using more than one Cop individual in the crosses with the F_2 as there is some variation in hybrid necrosis rates within the Cop population (Macnair & Christie, 1983; personal observations). Ideally, every cross would have used a single Cop parent but this was not feasible due to the number of flowers required to conduct the crosses and the relatively short flowering time of the Cop plant. Instead, five Cop individuals were used. Control crosses between a single Cer10 parent and two of the Cop parents gave an identical proportion of hybrid necrosis. Of the 153 plants that had replicates, seventy-two had the same Cop parent for both replicates whilst eighty-one had a different Cop parent.

As such, a total of 326 F_2 x Cop crosses from 168 F_2 individuals were sown with all but twelve having two replicates at the tray level. This gives a total number of 640 trays each containing thirty hybrid progeny scored for hybrid necrosis.

The progeny of the crosses were grown for between six to ten weeks then phenotyped for the number of green/healthy or yellow/necrotic/dead individuals to give the proportion of hybrid necrosis in the cross. The growing period was longer (approximately ten weeks) in the early spring months and shorter (approximately six weeks) in the late Spring/Summer so as to take into account the slower plant growth during the darker months.

6.2.4 Screening for markers

The markers tested for linkage to the QTLs for hybrid necrosis loci in the Cerig population were based on expressed sequence tags (ESTs). These were amplified across introns with the aim of identifying polymorphisms in intron length dependent on the different parent genomes of Cer10 and Cer35. The EST marker primers used were from the collection at Duke University and are described in 2.6.2.

As the QTL approach was targeted at *R* genes, the markers selected for screening were those located within, or near to, *R* gene clusters distributed throughout the *Mimulus* genome. To identify *R* gene clusters, the entire *Mimulus* genome was run through a blast search and the cluster size (number of *R* genes) and location were recorded. EST markers that were located within the same genomic region as an *R* gene cluster were then selected for screening. If the *R* gene cluster was large then more than one marker was used.

For *R* gene clusters where there were no near EST markers, primers were designed for new markers using mRNA (JGI annotation) sequences. The FASTA sequences of mRNAs within the region of the *R* gene cluster were extracted from GBrowse (www.mimulusevolution.org) and primers were designed for suitable introns (more than 150bp and flanked by exons/CDS of at least 20bp in which to design the primers). For large *R* gene clusters, markers were designed at 50kb intervals. The mRNA FASTA sequences with the target sequence defined, were put into an online programme, Primer3 (http://frodo.wi.mit.edu/primer3/), which designed pairs of primers. At the 5'end of each forward primer an additional fluorescently-labelled sequence (M13) was attached to enable cost-effective genotyping by the capillary sequencer.

DNA was extracted from Cer10, Cer35 and the F_2s primarily using the CTAB method and used in PCR reactions with the marker primers. A Qiagen DNeasy Kit was used to extract DNA from three F_2s where the CTAB extraction was unsuccessful. The PCR reactions were checked for successful amplification by gel electrophoresis. The PCR-amplified DNA fragments were sent for genotyping by capillary electrophoresis and fragment analysis on an ABI 3730x1 DNA Analyzer and analysed using GeneMarker.

In total, fifty EST markers and twenty-three new mRNA-based markers were screened in Cer10 and Cer35. Of the EST markers, five were selected due to their high polymorphism or reliable amplification rather than because they were within an *R* gene cluster region. The remaining forty-five EST markers were for *R* gene clusters ranging in size from four to forty-one *R* genes on twenty-four different scaffolds. The twenty-three new markers were for five scaffolds that contained large *R* gene clusters (between thirteen to forty-one *R* genes) where none, or very few, of the EST markers tested were polymorphic.

Of the seventy-three markers, one EST marker and five new markers failed to amplify a fragment in either one or both of the parents resulting in a total of sixty-eight markers that were genotyped in the Cer10 and the Cer35 parent plants to identify fragment length polymorphisms. Of these, twenty-five ESTs and two new markers were polymorphic and were subsequently used in BSA by testing them in the twelve F_2s that gave the highest necrosis and the twelve F_2s that gave the lowest necrosis. The frequency of marker genotypes in the high and low F_2 groups were analysed using Fisher's exact test. Those that were significant (i.e. were co-segregating with the hybrid necrosis phenotype) were then tested in the remaining F_2s that had not been used in the BSA. The positions of the markers tested are shown in figure 6.5.

6.2.5 Data Analysis Techniques

6.2.5.1 Fisher's exact test in BSA

Chi-squared can not be used when the expected values in any of the cells of the contingency table are less than five (Colquhoun, 1971). Therefore, as the modified BSA has a low sample number so that the values in the cells are frequently less than five, Fisher's exact test was used in preference to chi-squared to test for a significant relationship between marker genotype and hybrid necrosis. An online calculator (http://faculty.vassar.edu/lowry/fisher2x3.html) was used to perform Fisher's Exact Test as in a 2x3 table it becomes difficult to compute manually. The markers that were significant, and therefore whose genotype was linked to the hybrid necrosis phenotype, were subsequently tested in the remaining F_2s.

6.2.5.2 Sequential Bonferroni correction for multiple comparisons

To reduce Type I errors, a sequential Bonferroni correction was performed using the *p*-values obtained in 6.2.5.1. The *p*-values were put in ascending order and a new significance threshold calculated using the formula *0.5/k* for the smallest *p*-value, *0.5/k-1* for the second smallest *p*-value, *0.5/k-3* for the third smallest *p*-value and so on, where *k* is the number of comparisons. The *p*-values were significant if they were less than the new threshold value.

6.2.5.3 Arcsine transformation

The hybrid necrosis rate is usually given as the percentage of yellow/necrotic/dwarf/dead plants out of the total number of plants scored. For the purposes of statistical analysis using analysis of variance (ANOVA), the rate is given as a proportion (i.e. number of necrotic hybrids / total

number of hybrids scored). In order to use ANOVA the data must be homoscedastic so that the errors are normally distributed.

The residuals of the proportion of hybrid necrosis from the F_2 x Cop crosses were calculated for each F_2 by subtracting the mean proportion of hybrid necrosis from the individual F_2's necrosis rate. A histogram of these residuals, shown in figure 6.1, illustrates that the data are not normally distributed but have a strong positive skew. As the raw data of hybrid necrosis rates amongst the F_2 do not follow the normal distribution, they require a transformation step to normalise them in order to make them suitable for analysis by ANOVA.

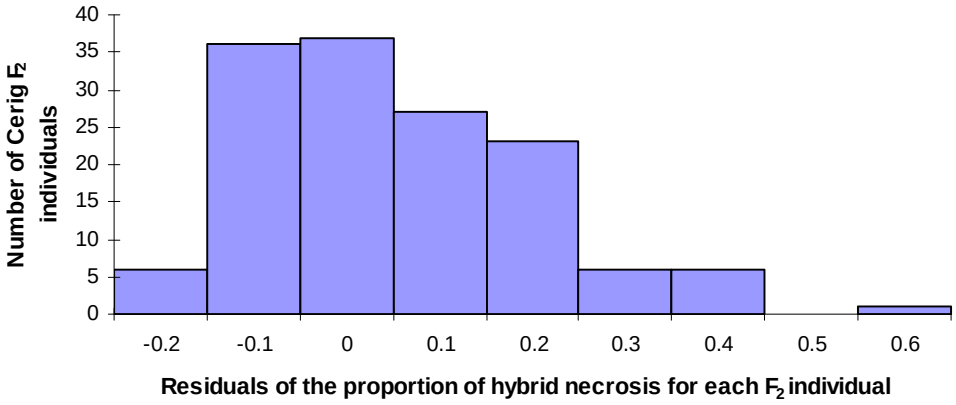


Figure 6.1 A histogram showing the frequency distribution of the residuals for the proportion of hybrid necrosis in crosses between Cop and 142 individuals from a segregating F_2 population of Cerig. The residuals were calculated using the formula: *proportion of hybrid necrosis of each F_2 individual - mean proportion of hybrid necrosis*.

Theory states that when working with proportions and binomially distributed data, the correct transformation to improve normality is arcsin. As such, the proportion of hybrid necrosis for each F_2 was arcsin-transformed by calculating the proportion of necrotic hybrids, finding its square root and then using the *[ASIN(square root P)*180/3.142]* formula in Microsoft Excel, and the residuals derived from this transformed data, shown in figure 6.2, are clearly much closer to the normal distribution than the residuals derived from the raw data thus proving that this transformation method has worked. The arcsin-transformed hybrid necrosis data, given in degrees, are in a format and distribution that can be analysed using ANOVA.

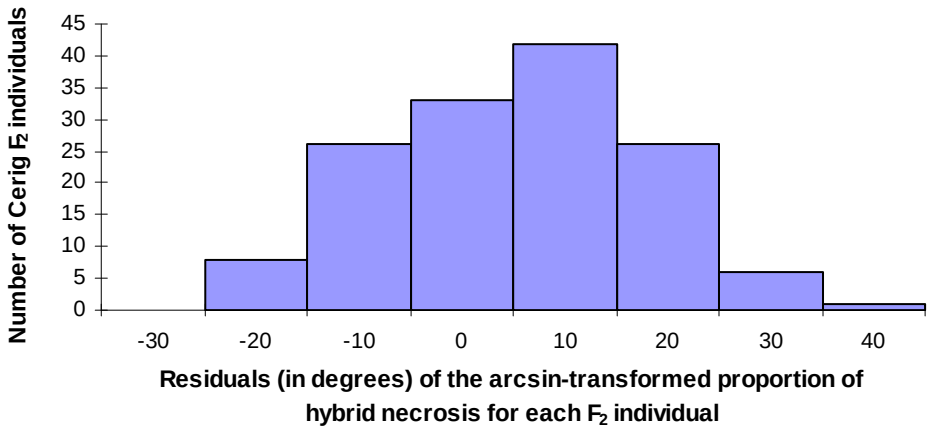


Figure 6.2 A histogram showing the frequency distribution of the residuals from the arcsin-transformed proportion of hybrid necrosis in crosses between Cop and 142 individuals from a segregating F_2 population of Cerig. The residuals were calculated using the formula: *arcsin-transformed proportion of hybrid necrosis of each F_2 individual - arcsin-transformed mean proportion of hybrid necrosis*.

6.2.5.4 One way Analysis of Variance (ANOVA)

For each marker that showed significance in Fishers exact test, a one-way, or single-factor, ANOVA was conducted using Excel to test for a difference between the hybrid necrosis means with each genotype. For each marker, the arcsin-transformed hybrid necrosis values with each of the three possible marker genotypes were used.

6.3 Results

6.3.1 Testing replicate F_2 x Cop crosses for homogeneity

At all levels of replication in the crosses between Cop and the segregating F_2 population, contingency chi-squared was used to test for homogeneity. The χ^2 values are given in Appendix 6.1.

At the tray level of replication, twenty-one of the 326 crosses had a significant difference ($p< 0.05$) in hybrid necrosis rates between the trays. Of these, ten are also significant at the $p<0.01$ level. Each of these twenty-one crosses involved a different F_2 parent meaning that none of the

F_2 individuals had significant differences between replicate trays in both of the crosses sown. These differences could therefore be due to type 1 errors in scoring necrosis or an effect of the position on the bench of the tray, for example the tray could be in a position that makes it more likely to incur a mildew infection that causes leaf necrosis that could be mistakenly scored as hybrid necrosis.

At the cross level, fifty-six crosses (and therefore fifty-six F_2 individuals) had a significant difference in their hybrid necrosis rate. Nine of these were the same F_2 that had a significant difference between the replicate trays. In thirty-six of the replicate crosses with the significant difference in the hybrid necrosis rates, a different Cop parent was used in the replicates raising the question of whether the Cop parent was significantly affecting the rate of necrosis. Given that a total of eighty-one F_2s had a different Cop parent for each replicate cross yet only thirty-six of these had a significant difference in the rate of necrosis and that there is no apparent trend in terms of the effect of the five Cop individuals on the necrosis rate, it is unlikely that different Cop parents could solely explain the difference in necrosis rates between replicates. The reasons for the apparent disparity in the hybrid necrosis rates between replicate crosses are probably due to errors in scoring necrosis and an effect of the growing environment.

As the aim was to establish the hybrid necrosis rate with each F_2, the data from all replicate trays and crosses for each F_2 were combined and also tested for homogeneity using chi-squared. Out of the 168 F_2s crossed with Cop, fifty had replicates whose hybrid necrosis rates were significantly different ($p<0.05$). Unexpectedly, eight of these fifty did not have significant differences between either the replicate trays or the replicate crosses implying that variation was greater in the larger sample size. Twenty-six had replicates that were significantly different at the $p<0.01$ level, two of which did not have significant differences between either the replicate trays or the replicate crosses.

Excluding individuals for which there was significant variation between trays provides a conservative baseline for excluding individuals showing significant environmental variation. However, it was important to use as many F_2s as possible in the QTL analysis and so the 142 F_2s that had no significant difference at the $p<0.01$ level were used. The hybrid necrosis proportions from all the replicate trays for each of the 142 F_2s were averaged in order to give one hybrid necrosis rate (percentage of hybrid necrosis) per F_2 individual. The remaining 26 F_2s whose replicates were significantly different at this level were excluded from further analysis as the difference in the hybrid necrosis between their replicates meant an average necrosis rate for these individual could not be reliably determined and could consequently affect the detection of linked markers. Two F_2 individuals ($F_2$3 and $F_2$6) were initially included despite their replicates having a large difference in the proportion of necrosis as their first replicate cross gave a zero

hybrid necrosis rate. These individuals were ultimately not included in any of the analyses and their initial inclusion and subsequent exclusion is described later.

6.3.2 Hybrid necrosis in the segregating F_2 population

The hybrid necrosis rate of a Cerig individual refers to the proportion of necrotic progeny that result from crossing it to Cop. The hybrid necrosis rates of the 142 individuals from the F_2 population of Cerig are shown in figure 6.3, from which it is evident that this is a segregating F_2 population as hybrid necrosis ranged from 0% to 76% with no distinct clustering. Figures 6.3 and 6.4 also reveal a large interval between the highest and second highest necrosis rates (75.83% and 59.06% respectively). There were in fact two additional hybrid necrosis rates in between these two values, but these were excluded from analysis as their replicates were significantly different at the *p<0.01* level.

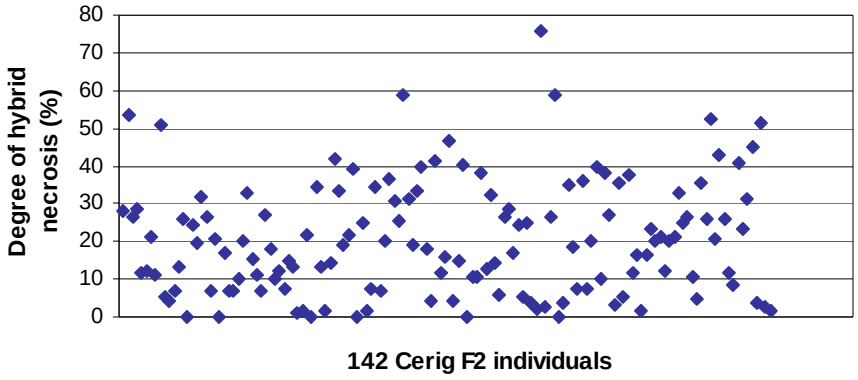


Figure 6.3 The wide variation in the distribution of the degree (proportion) of hybrid necrosis in crosses between Cop and 142 individuals from a segregating F_2 population of Cerig.

The wide range in necrosis rates is expected as there are a small number of polymorphic genes involved in the Cerig side of the crossing barrier (Macnair & Christie, 1983) and so the strength of the barrier (the proportion of necrotic hybrids) is likely to depend on the alleles at all these loci. However, although the distribution is wide, the majority (95%, or 135 out of 142 individuals) of the F_2s had hybrid necrosis rates of ≤50%, as shown clearly by the positive skew in the histogram in figure 6.4. This suggests that not all of the incompatible alleles required to cause the highest levels of necrosis are present in the F_2 population. This is due to the genotype of the Cer10 x Cer35 F_1 plant that was randomly selected and selfed to generate the F_2, an issue covered in greater depth in the discussion.

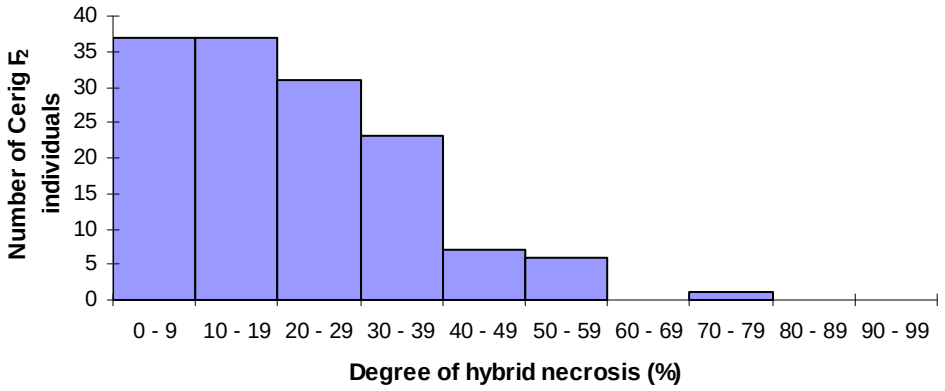


Figure 6.4 A histogram showing the frequency of the distribution of the degree (proportion) of hybrid necrosis in crosses between Cop and 142 individuals from a segregating F_2 population of Cerig.

6.3.3 Screening for polymorphic markers in the parent plants

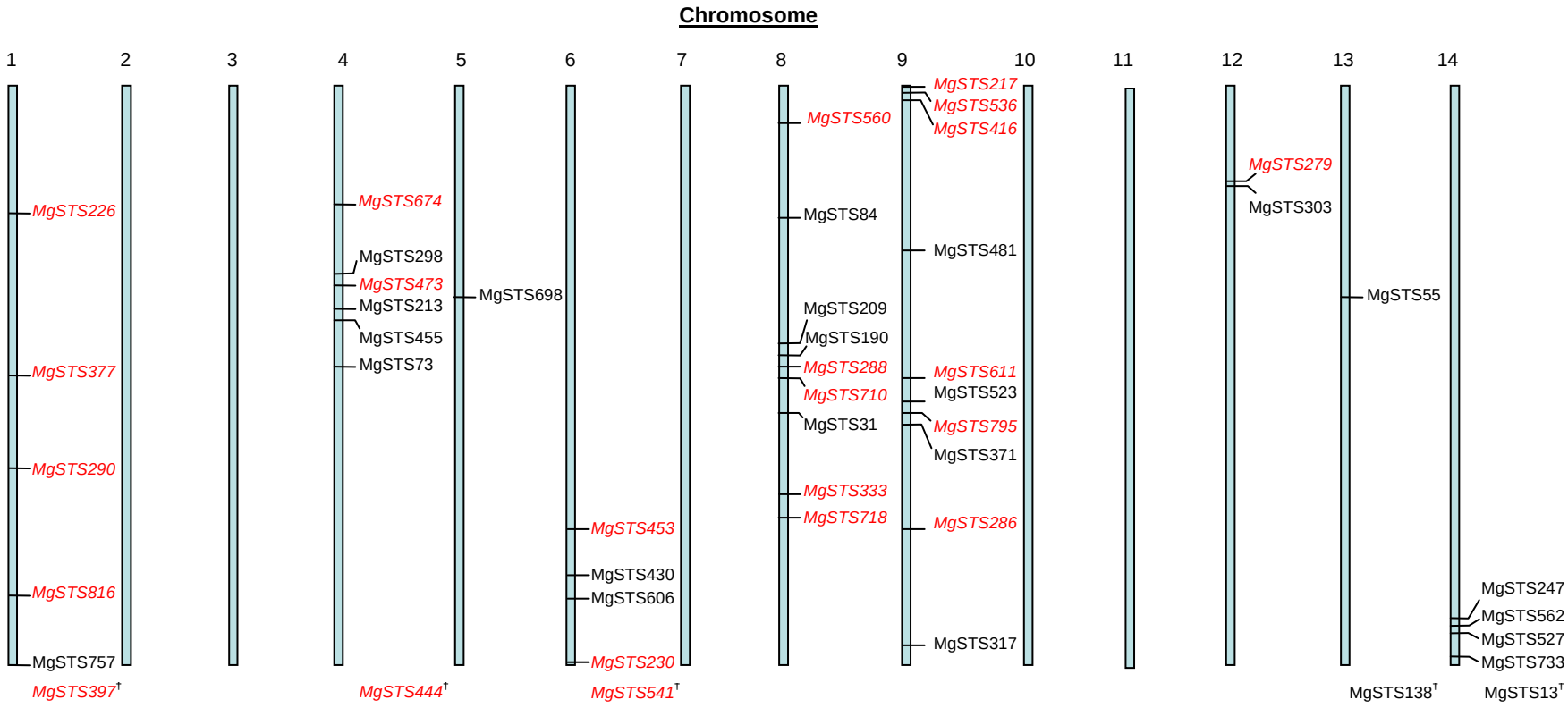
The chromosomal locations of the seventy-three markers genotyped in Cer10 and Cer35 are shown in figure 6.5. Of these markers, twenty-seven (indicated by red italics in figure 6.5) showed allele length polymorphism and their genotypes in the parent plants Cer10 and Cer35 are given in table 6.2. These markers, with the exception of e536 and sc96.735k which were omitted at this time due to cost limitations, were tested using the modified BSA in the twelve F_2s that gave the highest hybrid necrosis and the twelve F_2s that gave the lowest hybrid necrosis.

6.3.4 Bulked Segregant Analysis

The F_2 plants used in the BSA are shown in table 6.3. Due to an error in the initial analysis of the hybrid necrosis data for each F_2, some incorrect F_2s were used in the BSA. The explanation for the origin of the error and its implications are considered in greater detail in the discussion.

Table 6.2 The genotypes and location of the polymorphic markers identified in Cer10 and Cer35. Those shown in bold print were identified as co-segregating with the hybrid necrosis phenotype.

Marker	Scaffold no. and position on scaffold (base pairs)	Linkage Group	Cer10 genotype	Cer35 genotype
e217	**68: 778382 - 780439**	**9**	**202/205**	**205**
e226	79:1120959 - 1123783	1	270	253/270
e230	16: 475656 - 478155	6	129/196/198	196/198
e279	**132: 520125 - 521763**	**12**	**761/782**	**761**
e286	7: 1471694 - 1473014	9	350/351	353
e288	12: 1909367 - 1911522	8	536	516/536
e290	60: 1383430 - 1384716	1	341/342	356/357
e333	45: 1574435 - 1577311	8	222	218/222
e377	60: 117760 - 122264	1	435/436	436/437
e397	60: 621789 - 628074	1 (not mapped)	399	405
e416	**68: 201905- 203010**	**9**	**233/242**	**242**
e444	96: 144209 - 147806	4 (not mapped)	791	635/790
e453	16: 2567905 - 2568726	6	225/280	225
e473	96: 278012 - 279230	4	703	703/708
e485	14: 186889 - 189123	not mapped	407	409
e536	686: 250794 - 252493	9	255/259	255
e541	16: 2131416 - 2132848	6 (not mapped)	794	799
e560	70: 121599 - 434469	8	208/240	208
e611	7: 3311584-3313523	9	176	171/176
e674	85: 380140 - 380838	4	567	506/567
e707	95: 554834 - 555509	not mapped	229/231	228/229
e710	12: 2675505 - 2677683	8	414	333/414
e718	45: 923708 - 927040	8	406	416/418
e795	7: 2738477 - 3742186	9	258	253/258
e816	240: 70825 - 79577	1	205/206	207
sc96.735k	96: 735540 - 739030	4 (not mapped)	685	695
sc96.825	96: 825947 - 830523	4 (not mapped)	230/241	230



† Marker not mapped but located on the same scaffold as other markers that have been mapped to this chromosome

Other markers screened but map position not known:

EST Markers: *MgSTS485*; MgSTS732; *MgSTS707*

New Markers: Scaf6-DL_189k; Scaf6-DL_265k; Scaf6-DL_455k; Scaf6-DL_598k; Scaf6-DL_682k; Scaf6-DL_848k; Scaf55-DL_397k; Scaf55-DL_507k; Scaf55-DL_833k; Scaf55-DL_988k; Scaf96-DL_295k; Scaf96-DL_515k; Scaf96-DL_655k; *Scaf96-DL_735k*; *Scaf96-DL_825k*; Scaf57-DL_698k; Scaf57-DL_786k; Scaf57-DL_866k; Scaf57-DL_969k; Scaf14-DL_392k; Scaf14-DL_451k; Scaf14-DL_627k; Scaf14-DL_209k

Figure 6.5 The relative positions on the 14 linkage groups of the *M. guttatus* genome of the seventy-three markers screened for polymorphism in Cer10 and Cer35 (from map information available at www.mimulusevolution.org). The twenty-seven markers that were polymorphic in plants A and D are shown in red italics

Within the high-necrosis group, two individuals that caused 52.25% and 41.46% hybrid necrosis were not used in the BSA. Instead, two that caused 25.98% and 25.55% were used whose hybrid necrosis rates, although above the average rate of 20.44%, fell mid-range (rank 48 and 50 out of 137, where rank 137 = 0% necrosis) rather than at the high extreme. In the low-necrosis group, five of the lowest necrosis F_2 were incorrectly omitted and instead three F_2s of slightly higher necrosis (1.92% rank 130, 3.96% rank 124 and 11.12% rank 98, where rank 137=0%) were used as well as two individuals, $F_2$3 and $F_2$6, that were mistakenly used due to the highly different hybrid necrosis proportions between the replicates at the cross level. With $F_2$3, the hybrid necrosis in the first replicate cross was 26.67% and as such was significantly different from the 0% necrosis in the second replicate (χ^2 = 18.462 significant at *p<0.001, 1df*). Similarly, with $F_2$6 the first replicate had a hybrid necrosis rate of 21.67% but the second had 0% (χ^2 = 14.579 significant at *p<0.001, 1df*). Upon discovery of the error in the assignment of their hybrid necrosis rates, and given their genotypes in BSA, these two plants were omitted from all further analysis including the Fisher's exact tests on the BSA results so that the low hybrid necrosis group contained ten instead of twelve individuals.

Table 6.3 The hybrid necrosis rates of the F_2 plants that were selected for BSA and of those plants which should have been selected.

† = F_2 incorrectly used in BSA; †† = F_2 incorrectly used in BSA and has a significant difference between replicate crosses so is excluded from use in further analysis; ††† = F_2 which should have been used in BSA but was not.

F_2s used in BSA				**F_2s that should have been used in BSA**			
High Necrosis Group	**Hybrid Necrosis rate (%)**	**Low Necrosis Group**	**Hybrid Necrosis rate (%)**	**High Necrosis Group**	**Hybrid Necrosis rate (%)**	**Low Necrosis Group**	**Hybrid Necrosis rate (%)**
145	75.833	22	0.000	145	75.833	22	0.000
105	59.065	44	0.000	105	59.065	44	0.000
148	58.851	93	0.000	148	58.851	80	0.000
4	53.333	121	0.000	4	53.333	93	0.000
201	51.333	149	0.000	189†††	52.248	121	0.000
14	50.833	76	0.833	201	51.333	149	0.000
116	46.381	78	1.667	14	50.833	76	0.833
199	44.951	142†	1.923	116	46.381	78	1.667
191	42.655	150†	3.960	199	44.951	83†††	1.667
88	41.667	6††	21.667	191	42.655	204†††	1.667
188†	25.977	63†	11.121	88	41.667	170†††	1.695
104†	25.555	3††	26.270	112†††	41.459	95†††	1.724

The errors in the selection of the plants used in BSA were discovered subsequent to completion of the BSA marker testing in the F_2. Fisher's exact tests were conducted to test whether the marker genotypes were associated with the hybrid necrosis phenotype, so in order to establish what implications the use of the incorrect plants had on the results this statistical analysis was performed twice: once with all the plants used in the BSA (minus $F_2$3 and $F_2$6 from the low group) and once with only those plants that were correct (i.e. omitting those plants marked with † in table 6.3). The results of Fisher's exact tests (given in table 6.4) indicate that although the use of the incorrect plants do affect the p-value, with the exception of one marker (e279) their use does not alter whether the markers are found to be significant in co-segregating with the hybrid necrosis trait. As such, the results from the analysis of the data from all the plants used in BSA (except $F_2$3 and $F_2$6) were usable.

The frequency of the genotypes for the twenty-five polymorphic markers tested in the high and low necrosis groups and the values obtained in Fisher's exact tests shown in table 6.4 reveals three markers (*) with genotypes that co-segregate with a QTL controlling the hybrid necrosis phenotype: e217, e416 and e279. The inclusion of marker e279 in the subsequent analysis, despite its proximity to the statistical threshold and its non-significance once a sequential Bonferroni calculation for multiple comparisons has been performed, is considered in the discussion.

6.3.5 Marker and QTL analysis

The three markers, e217, e416 and e279, identified by BSA as co-segregating with the necrosis phenotype were tested in the remaining individuals from the segregating F_2 population of Cerig.

The p-values, shown in table 6.5, obtained by conducting one way ANOVAs on each marker are all significant (<0.05), meaning that with each marker the mean hybrid necrosis for each of the three genotypes were significantly different. As such, these markers co-segregate with the hybrid necrosis phenotype and thus are linked to QTLs harbouring genes responsible for hybrid necrosis. The full ANOVA table from Excel for each marker is given in Appendix 6.2.

Table 6.4 The frequency of the genotypes in the BSA high- and low- necrosis F_2 groups of the twenty-five polymorphic markers in Cer10 and Cer35. The three markers likely to co-segregate with hybrid necrosis based on the significant values from Fisher's exact test are shown in bold italics.

** = significant in Fisher's exact test (p<0.05); ** = not significant (p>0.05) in Fisher's exact test with correct plants only but significant (p<0.05) with Fisher's exact test on all plants used in BSA; BT = significance threshold from sequential Bonferroni correction for multiple comparisons; † = significant (p<BT) after sequential Bonferroni correction.*

Marker	Genotype			Frequency of genotype				Fisher's exact on F_2 used in BSA	Fisher's exact on correct F_2 only
				High necrosis group		Low necrosis group			
	Cer10	Cer35	Hybrid	F_2 used in BSA (12 plants)	Correct F_2 only (10 plants)	F_2 used in BSA (10 plants)	Correct F_2 only (7 plants)		
e217*	***202/205***	***205***	***202/202***	***1***	***1***	***0***	***0***	***0.000704**** ***BT = 0.0294†***	***0.00339**** ***BT = 0.0294†***
			202/205	***11***	***9***	***3***	***2***		
			205/205	***0***	***0***	***7***	***5***		
e226	270	253/270	270/270	3	2	4	2	0.587 BT = 0.0455	1.000 BT = 0.111
			270/253	7	6	4	4		
			253/253	1	1	2	1		
e230	128/129/ 196/198	196/198	128/196	1	1	1	1	0.185 BT = 0.0357	0.338 BT = 0.0357
			129/198	6	6	8	6		
			128/129/ 196/198	5	3	1	0		
e279*	***761/782***	***761***	***761/761***	***4***	***4***	***8***	***6***	***0.0427**** *BT = 0.0333*	***0.134***** *BT = 0.0333*
			761/782	***8***	***6***	***2***	***1***		
			782/782	***0***	***0***	***0***	***0***		
e286	351	353	351/351	9	8	7	5	1.000 BT = 0.143	1.000 BT = 0.111
			351/353	3	2	3	2		
			353/353	0	0	0	0		
e288	536	516/536	536/536	10	9	7	5	0.624 BT = 0.0526	0.537 BT = 0.0476
			536/516	2	1	3	2		
			516/516	0	0	0	0		
e290	341/342	356/357	341/341	12	10	10	7	N/A	N/A
e333	222	218/222	222/222	10	9	7	5	0.624 BT = 0.0526	0.537 BT = 0.0476
			222/218	2	1	3	2		
			218/218	0	0	0	0		

Table 6.4 *continued*

Marker	Genotype			Frequency of genotype				Fisher's exact on F_2 used in BSA	Fisher's exact on correct F_2 only
				High necrosis group		Low necrosis group			
	Cer10	Cer35	Hybrid	F_2 used in BSA (12 plants)	Correct F_2 plants only (10 plants)	F_2 used in BSA (10 plants)	Correct F_2 plants only (7 plants)		
e377	435/436	436/437	435/435	12	10	10	7	N/A	N/A
e397	399	405	399/399 399/405 405/405	7 3 2	7 2 1	6 3 1	4 2 1	1.000 BT = 0.143	1.000 BT = 0.111
e416*	***233/242***	***242***	***233/233*** ***233/242*** ***242/242***	***1*** ***11*** ***0***	***1*** ***9*** ***0***	***0*** ***2*** ***8***	***0*** ***1*** ***6***	***0.000141**** ***BT= 0.0313*** †	***0.000565**** ***BT= 0.0313*** †
e444	791	635/790	635/635	12	10	10	7	N/A	N/A
e453	225/280	225	225/225 225/280 280/280	- - 4	- - 2	- - 2	- - 1	-	-
e473	703	703/708	703/703 703/708 708/708	8 4 0	6 4 0	3 5 1	3 3 1	0.255 BT = 0.0417	0.773 BT = 0.0556
e485	407	409	407/407 407/409 409/409	0 12 0	0 10 0	0 8 2	0 6 1	0.195 BT = 0.0385	0.412 BT = 0.0385
e541	794	799	No amp	-	-	-	-	-	-
e560	240/208	208	208/208	12	10	10	7	N/A	N/A
e611	176	171/176	176/176 176/171 171/171	- - 3	- - 2	- - 3	- - 2	-	-

Table 6.4 *continued*

Marker	Genotype Cer10	Genotype Cer35	Hybrid	Frequency of genotype: High necrosis group F_2 used in BSA (12 plants)	High necrosis group Correct F_2 plants only (10 plants)	Low necrosis group F_2 used in BSA (10 plants)	Low necrosis group Correct F_2 plants only (7 plants)	Fisher's exact on F_2 used in BSA	Fisher's exact on correct F_2 only
e674	567	506/567	567/567	7	5	4	4	0.802 BT = 0.0714	0.471 BT = 0.0417
			567/506	5	5	4	2		
			506/506	0	0	1	1		
e707	229/231	228/229	229/229	3	3	2	1	1.000 BT = 0.143	1.000 BT = 0.111
			229/228	1	1	0	0		
			231/229	7	5	8	6		
			231/228	0	0	0	0		
e710	414	333/414	414/414	9	8	6	5	0.652 BT = 0.0625	1.000 BT = 0.111
			414/333	3	2	4	2		
			333/333	0	0	0	0		
e718	405	417	405/405	1	1	0	0	1.000 BT = 0.143	1.000 BT = 0.111
			405/417	9	7	5	3		
			417/417	0	0	0	0		
e795	258	253/258	258/258	9	8	7	5	1.000 BT = 0.143	1.000 BT = 0.111
			258/253	3	2	3	2		
			253/253	0	0	0	0		
e816	205/206	207	206/206	12	10	10	7	1.000 BT = 0.143	1.000 BT = 0.111
			206/207	0	0	0	0		
			207/207	0	0	0	0		
Sc 96. 825	230/241	230	No PCR amplification	-	-	-	-	-	-

Table 6.5 The genotypes and one-way ANOVA p-values for the three markers identified for QTLs controlling hybrid necrosis in Cerig. All p-values are <0.05 and are therefore significant.

Marker	Cer10 genotype (high necrosis)	Cer35 genotype (low necrosis)	F_2 Genotype	Number of plants with genotype	Average Necrosis %	ANOVA p-value
e217	202/205	205	202/202	8	24.64	1.08×10^{-16}
			202/205	79	27.77	
			205/205	51	8.60	
e416	233/242	242	233/233	11	25.17	2.04×10^{-20}
			233/242	76	28.34	
			242/242	50	7.92	
e279	761/782	761	782/782	0	-	0.024539
			761/782	30	27.81	
			761/761	79	19.39	

Markers e217 and e416: Linkage Analysis

Markers e217 and e416 are both located on scaffold_68 at 201906 - 203010bp and 778381 - 780345bp respectively and map to linkage group (LG) 9. Interestingly, this is the same chromosome that, in Copperopolis, harbours both the single locus, *NEC,* which complements the DM-genes in Cerig to cause necrosis and the copper tolerance locus, *T*, although they appear to be some distance apart. The map locations of these loci are shown in figure 6.6. Given their map positions, it is not surprising that the results from these markers e217 and e416 are very similar, as shown in figure 6.7, so determining the linkage between these markers could reveal information about their linkage to necrosis.

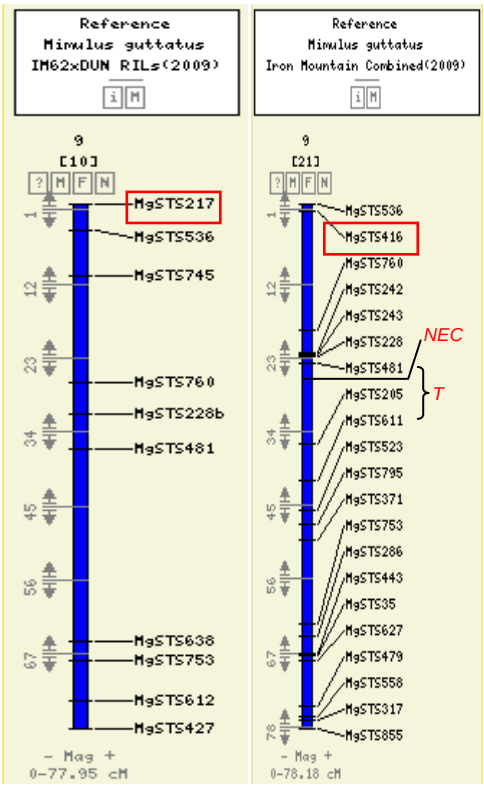


Figure 6.6 Maps of Chromosome 9 of the *M. guttatus* genome showing the location of markers e217 and e416 along with the location of NEC and T loci as identified by Chapters 4 & 5 (map images taken from www.mimulusevolution.org).

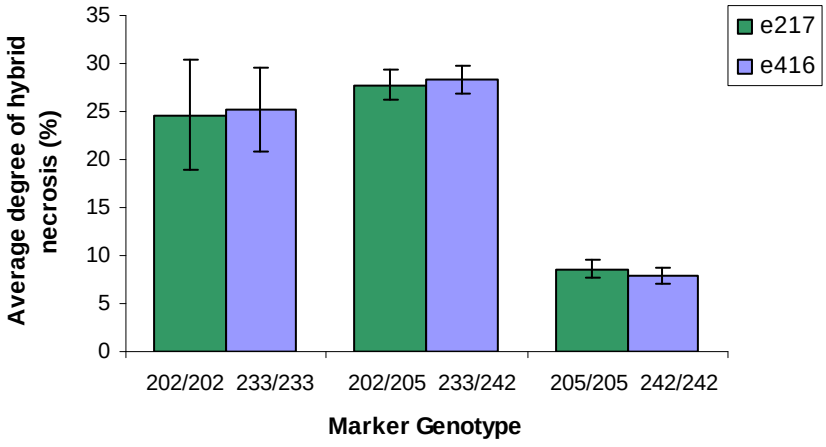


Figure 6.7 The average degree of hybrid necrosis associated with each of the three genotypes of markers e217 and e416 in the segregating F_2 population of Cerig.

The frequency of the nine possible genotypes from the combination of e217 and e416 are given in table 6.6. The two alleles of marker e217, *202* and *205*, are denoted as *A* and *a* respectively

whilst the two alleles of marker e416, *233* and *242*, are denoted as *B* and *b* respectively. Table 6.2 showed that Cer10 had the genotype *AaBb* and so must have provided the *AB* chromosome whilst Cer35 had the genotype *aabb* and thus provided the *ab* chromosome. The parental (P) chromosomes are therefore *AB* or *ab*, any other combination (*aB* or *Ab*) is a recombinant (R) chromosome.

Table 6.6 The frequency of the nine possible genotypes from the combination of markers e217 and e416 in the 137 F_2 hybrids with genotypes at these two markers and the average degree of hybrid necrosis (%) associated with each genotype.

		e217			**Total no. of F_2**
		202/202 (AA)	*202/205 (Aa)*	*205/205 (aa)*	
e416	*233/233 (BB)*	8 (24.64%)	3 (26.60%)	0	11
	233/242 (Bb)	0	74 (28.51%)	2 (22.03 %)	76
	242/242 (bb)	0	2 (2.20%)	48 (8.15%)	50
Total no. of F_2		8	79	50	137

There are a total of 137 plants in which the genotype at both e217 and e416 are known. This is equivalent to 274 chromosomes. There are seven F_2s in which recombination between e217 and e416 must have occurred during meiosis in the F_1 individual that was selfed to produce the F_2 as the genotypes *A**a**B**B*** (three individuals), ***A**a**b**b*** (two individuals), and ***a**a**B**b*** (two individuals) are represented in the F_2. The recombination frequency (RF) between linked genes can therefore be calculated using the frequency of P and R chromosomes shown in table 6.7.

Table 6.7 The number of parental (P) and recombinant (R) chromosomes present in the Cerig F_2 population in the combined genotypes of markers e217 and e416.

Genotype	**Number of F_2**	**Number of Parental chromosomes in genotype**	**Number of Parental chromosomes in F_2 population**	**Number of Recombinant chromosomes in genotype**	**Number of Recombinant chromosomes in F_2 population**
AABB	8	2 (AB)	16	0	0
AaBB	3	1 (AB)	3	1 (aB)	3
aaBB	0	0	0	2 (aB)	0
AABb	0	1 (AB)	0	1(Ab)	0
AAbb	0	0	0	2 (Ab)	0
AaBb	74	?	?	?	?
Aabb	2	1 (ab)	2	1 (aB)	2
aaBb	2	1 (ab)	2	1 (aB)	2
aabb	48	2 (ab)	96	0	0
Total	**137**	8	**119**	8	**7**

For the double heterozygotes *(AaBb*), it is not known how many possess two parent chromosomes (PP) (*AB/ab*) or how many are double recombinants (RR) (*Ab/aB*). If they are all PP, the recombination rate between *A* and *B* is 2.5% (7/274 = 0.025). Given this, the expected number of RR in the double heterozygotes would be 0.04 plants (0.025 x 0.025 x 74). Even if one of the double heterozygotes is a double recombinant, the recombination rate is still low at 3.3% (9/274 = 0.033) meaning it is still improbable that there are any double heterozygotes which are RR. Thus, the most likely estimate of the RF is 2.5cM, showing that recombination between e217 and e416 is rare as is expected due to their close physical proximity.

However, when recombination between these two marker loci does occur it may affect their linkage to the necrosis phenotype depending on the order of the loci along the chromosome. The results of testing this by comparing the necrosis rates of each of the three recombinant genotypes to the rate in the nearest unrecombined genotype are given in table 6.8.

There are two chi-squared values that stand out as being highly significant and both result from comparing a recombinant genotype to a non-recombinant genotype in which the allele at the *B* locus (i.e. marker e416) has altered. When the *aaBb* recombinant is compared to the non-recombinant *aabb*, the absence of the *B* allele has a highly significant impact on the rate of necrosis (χ^2 = 62.086, 1df) as it is associated with a decrease in hybrid necrosis from 22.03% to 8.15%. Similarly, comparison of the recombinant *Aabb* with the non-recombinant *AaBb* shows that necrosis increases 13-fold from 2.20% to 28.51 % respectively (χ^2 =77.098, 1df) with the presence of the *B* allele.

Furthermore, when the necrosis rate of the recombinant *aaBb* is compared to the non-recombinant *AaBb*, although there is some difference in the rates (χ^2 = 4.807, 1df) it is minor in comparison to the chi-squared values described above and could be attributed to the small sample size of the recombinant genotypes. As such, this indicates that the loss of the *A* allele during recombination does not affect the rate of necrosis suggesting that recombination between e217 and e416 breaks the linkage between the *A* allele of marker e217 and the necrosis phenotype. When only the *A* allele is present (*Aabb*) necrosis is low (2.20%) and, in comparison to other chi-squared values in the table, only slightly different (χ^2 = 8.712, 1df) from the compatible allele homozygote *aabb*, whose necrosis is actually higher (8.15%).

Thus, comparing the necrosis rates of these genotypes suggests that the recombination event between e217 and e416 causes dissociation of the linkage between marker e217 and hybrid necrosis whilst marker e416 remains linked to necrosis.

Table 6.8 The chi-squared (χ^2) values from comparing the degree of hybrid necrosis associated with the three recombinant genotypes of the markers e217 and e416 with the degree associated with the nearest non-recombinant genotypes. The genotypes that are highly different in the degree of necrosis with which they are associated are shown in bold.

*(significant at 1df *p<0.05, **p< 0.01, ***p< 0.001)*

Recombinant genotype	Necrosis rate %	Unrecombined (parental) genotype	Necrosis rate %	χ^2
AaBB (AB/aB)	26.60	*AABB (AB/AB)*	24.64	11.670**
		AaBb (AB/ab)	28.51	4.885*
aaBb (aB/ab)	22.03	*AaBb (AB/ab)*	28.51	4.807*
		aabb (ab/ab)	8.15	**62.086*** **
Aabb (Ab/ab)	2.20	***AaBb (AB/ab)***	28.51	**77.098*** **
		aabb (ab/ab)	8.15	8.712**

Although all significant, the other chi-squared values in table 6.8 are relatively small and so are probably due to the effects of the small sample size of the recombinant F_2 genotypes (three F_2 with the AaBB genotype, two with the aaBb genotype and two with the Aabb genotype; see table 6.6) than the effects of the different genotypes. However, it should be kept in mind that with such a small sample size of recombinant plants the reliability of the data could be affected.

So, this analysis reveals that markers e217 and e416 are very tightly linked with a recombination frequency of just ~2.5%. Recombination between these markers, although infrequent, causes a break in the association of e217 with the necrosis phenotype suggesting that e416 is more tightly linked to the necrosis locus. Furthermore, as this pattern is consistent across all the recombinants, the true gene order along chromosome 9 is most likely e217 – e416 – DM-gene locus.

Given their map location and the fact that the rate of recombination increases with genetic distance between loci and given that there is little recombination, these two markers are for a single QTL on Chromosome 9 that is associated with the hybrid necrosis trait. As the necrosis phenotype consistently segregates with the *B* allele (i.e. the incompatible allele of marker e416) it is reasonable to describe further analysis with reference only to marker e416, especially as the results of marker e217 follow exactly those of e416.

Marker e416

Marker e416 had the following three genotypes: *233/233* (eleven individuals), *233/242* (seventy-six individuals) and *242/242* (fifty individuals). PCR amplification failed to produce a fragment in five F_2 individuals, four of which also had no amplification with marker e217. Figure 6.8 shows the average necrosis rate associated with each genotype of e416 and illustrates the clear and significant decrease in hybrid necrosis when the *B* (*233)* allele is absent.

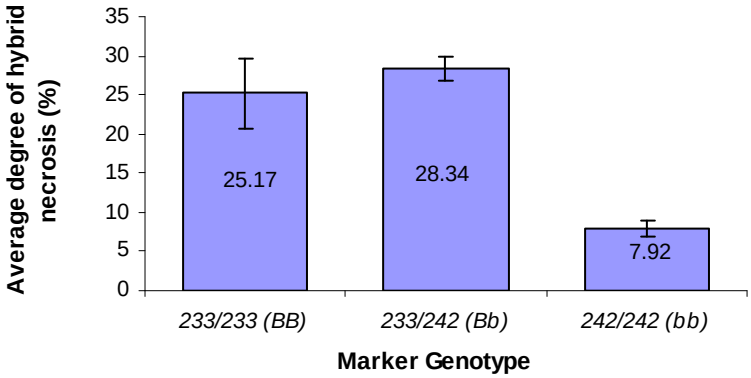


Figure 6.8 The average degree of hybrid necrosis associated with the three genotypes of marker e416 in the segregating F_2 Cerig population.

From the ratios of F_2 genotypes it is very evident that there is segregation distortion (SD), which is common in *Mimulus,* causing a large deviation away from the expected Mendelian proportion of 1:2:1 individuals in each genotypic class within the segregating F_2 population. With the 137 F_2 individuals that gave a genotype with this marker, the expected genotype ratio is 34.25 *BB* : 68.5 *Bb* : 34.25 *bb* which is significantly different (χ^2 = 836.375, p<0.001, 2df) from the observed ratios of 11:76:50. The observed ratio of genotypes suggests that SD distorts in favour of the *b* allele.

The *B* allele comes from the high-necrosis conferring Cer10 parent and is linked to a 'necrosis' allele that is incompatible with *NEC* in Cop whilst the *242* or *b* allele can be contributed either by Cer10 or by Cer35 and is linked to a compatible allele. The difference between the non-necrosis genotype (*bb*) and the necrosis genotypes (*BB* and *Bb*) is 18.84% meaning the DM-gene marked by this marker accounts for ~20% of the hybrid necrosis in the cross between Cop and Cerig.

Unexpectedly, the standard error bars overlap between the homozygous necrosis genotype *BB* and the heterozygous *Bb* genotype indicating that the *B* allele is associated with the same average degree of hybrid necrosis whether the plant is homozygous or heterozygous for this allele. This is confirmed by chi-squared which shows the difference between the hybrid necrosis rate with these two genotypes is not significant (χ^2 = 2.518 not significant at *p<0.05, 1df*). Given that the progeny of a *BB* parent will all receive the incompatible allele whilst, under independent assortment, only half will receive an incompatible allele from a *Bb* parent, it was expected that the necrosis rate in the latter should be half that of the necrosis homozygote. However, the distribution of hybrid necrosis rates with the three genotypes of marker e416 as shown in figure 6.9 indicates that the necrosis rate of *Bb* does lie in between *BB* and *bb*, as expected. The possible reasons for the similar average necrosis rates with *BB* and *Bb* are considered in the discussion.

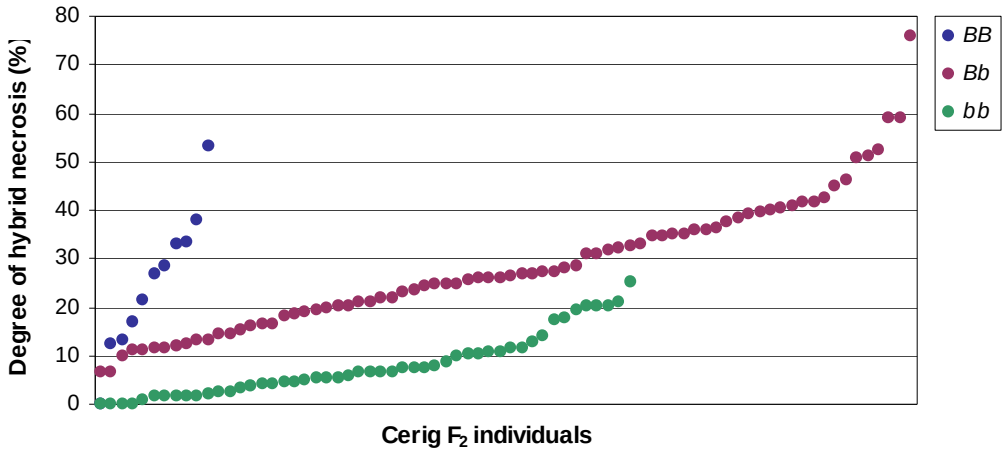


Figure 6.9 The distribution of the degree of hybrid necrosis associated with each individual in the segregating F_2 population with respect to their genotype at marker e416.

Marker e279

Marker e279 lies at 520126-521716kb on scaffold_132 which maps to chromosome 12 as shown in figure 6.10. The ANOVA result is significant *(<0.05*) meaning that this is a marker for a QTL underlying one of the loci responsible for the genic incompatibility that leads to hybrid necrosis.

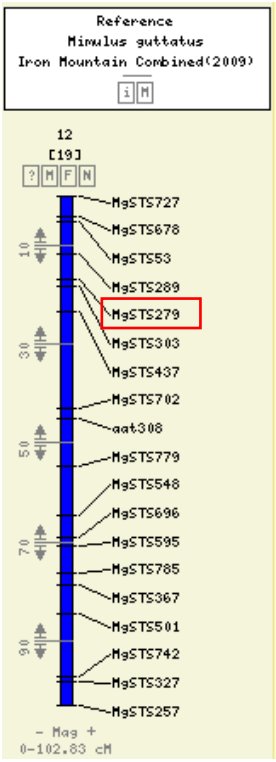


Figure 6.10 A map of chromosome 12 of the *M. guttatus* genome showing the location of marker e279 (image taken from www.mimulusevolution.org)

In the F_2 population, marker e279 had the following two genotypes: *761/761* (seventy-nine individuals) and *761/782* (thirty individuals). The *782/782* genotype did not appear and thirty-three individuals were not genotyped for this marker due to failed PCR amplification. These included the five without a genotype for e416. The average hybrid necrosis rates associated with each genotype of e279 are given in figure 6.11 and show the significant increase in hybrid necrosis that is associated with the *782* allele.

The *782* (*C*) allele is contributed by the high-necrosis conferring Cer10 parent whilst the *761* (*c*) allele can be contributed either by Cer10 or by Cer35. As such, the *C* allele of the marker is associated with the 'necrosis' allele that is incompatible with *NEC* in Cop, whilst the *c* allele is associated with the compatible allele.

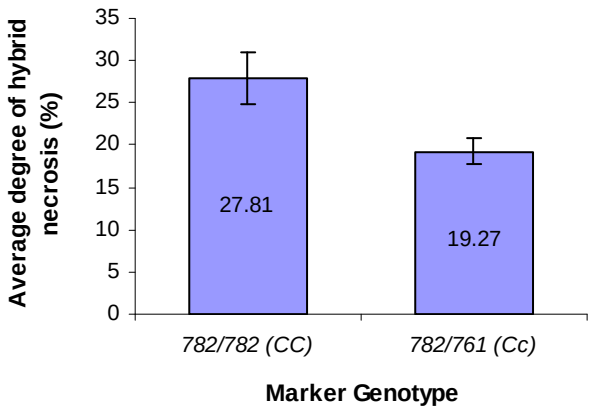


Figure 6.11 The average degree of hybrid necrosis associated with the two genotypes of marker e279 present in the segregating F_2 population of Cerig.

Without the presence of the *CC* genotype, it is hard to establish for certain whether there is segregation distortion at this locus, but the ratio of genotypes for *CC*:*Cc*:*cc* in the F_2 is 0:30:79 which is significantly different (χ^2 = 114.36 significant at p<0.001, 2df) from the expected ratio under Mendelian inheritance of 34.25:68.5:34.2 thus indicating segregation distortion in favour of the *c* allele.

At 19.27%, the *cc* genotype is associated with unexpectedly high necrosis considering it is linked with the compatible allele of the DM-gene and should not therefore be associated with necrosis. This relatively high necrosis rate could be due to the presence in the same F_2 genome of incompatible alleles of other DM-genes which have an effect on the rate of necrosis and so cause this marker to appear associated with a misleading high necrosis rate. The effect of other incompatible alleles on the necrosis caused by an incompatible allele at one locus can be seen when this marker is considered alongside e416.

The combined results of markers e279 and e416

Out of the nine possible genotypes with these two markers, there are five genotypes represented in the F_2 population. A total of 109 F_2s out of the 142 tested had a genotype for both markers. The number of F_2s with each genotype and the average necrosis rates associated with them are shown in table 6.9.

Table 6.9 The five combined genotypes of markers e416 and e279 present in the Cerig F_2 population and their associated necrosis rates.

Genotype at e416	Genotype at e279	Number of F_2	Necrosis rate
233/233 (BB)	*761/761 (cc)*	6	25.66
233/242 (Bb)	*782/761 (Cc)*	22	35.96
233/242 (Bb)	*761/761 (cc)*	47	24.92
242/242 (bb)	*782/761 (Cc)*	8	5.39
242/242 (bb)	*761/761 (cc)*	26	7.57

As the necrosis rates of all genotypes including *B* are all near to or in excess of 25% but below 10% when this allele is absent, a statistically significant difference (ANOVA: *F=103.144, P<0.0001*), it indicates that the *B* allele is consistently linked to necrosis. In addition, the presence of *B* is necessary for *C* to have an effect on the necrosis rate, as in the two *bb* homozygotes the presence or absence of the *C* allele does not significantly affect the necrosis rate (ANOVA: *F= 0.466, P>0.1*) with a difference between the two relevant genotypes of only ~2%. However, the necrosis rate increases significantly from 24.92% to 35.96% (ANOVA: *F = 12.527, P = < 0.001*) with the presence of the *C* allele in addition to the *B* allele indicating that *C* acts as an enhancer of *B*.

So, after deducting the background rate of necrosis (that associated with the *bbCc* and *bbcc* genotypes), the hybrid necrosis associated with the *B* allele of marker e416 is around 20% which is increased to around 30% by the additive effect of necrosis associated with the *C* allele of marker e279.

6.4 Discussion

This first investigation into the small number of DM-genes in Cerig which interact with the Copperopolis *NEC* locus to cause hybrid necrosis has identified two QTLs for hybrid necrosis, one on chromosome 9 and the other on chromosome 12.

On chromosome 9, linkage analysis reveals very low recombination between the necrosis markers e217 and e416 (RF of 2.5%) with all recombinants showing the same pattern where the necrosis phenotype remains associated with marker e416. As such, the order of these genes is most likely e217 – e416 – DM-gene locus, with the gene(s) marked by e416 being responsible for causing ~20% of the hybrid necrosis. The QTL on chromosome 12 accounts for ~10% necrosis, but its effects are only found when the incompatible allele on chromosome 9 is

also present. The effects of the incompatible alleles on these two QTL appears to be additive, as together they account for around 30% of the necrosis. As such, the DM gene on chromosome 12 acts as an enhancer of the DM gene on chromosome 9.

It is important to acknowledge that in the BSA the significance of marker e279, which identified the QTL on chromosome 12, was close to the statistical threshold and in fact it became non-significant once a sequential Bonferroni calculation for multiple comparisons had been performed. This would suggest that this marker, and therefore QTL, is not strongly associated with necrosis. Nevertheless, as only two other markers were identified as being linked to hybrid necrosis, it was considered worthwhile to include this marker in subsequent analysis. The one-way ANOVA that was performed using all the plants genotyped at marker e279 was significant indicating that, despite the lack of significance in the BSA analysis, there is in fact an association between e279 and hybrid necrosis. Perhaps the inconsistency in the statistical significance of this marker is a reflection of nature of the locus on chromosome 12, which appears to have an affect on necrosis by enhancing the degree of necrosis caused by the locus on chromosome 9, and as such the genotype at marker e416 will strongly affect the association of marker e279 to necrosis which, in turn, is likely to affect the statistical significance of this marker.

Rather than using random markers to identify the QTLs, the markers used were for *R* gene clusters distributed throughout the *Mimulus* genome. Marker e416 is located on scaffold_68, at 201906 - 203010bp along with seven *R* genes positioned in five distinct groupings (6893-7975bp, 70503-72558bp, 230618-232867bp, 366968-368434bp, and 668797-775917bp). Marker e279 lies at 520126-521716kb on scaffold_132 which contains eight *R* genes at two clusters, one at 189100-271364bp and one at 518587-531698bp. As such, the QTLs identified by these two markers are likely to contain *R* genes, meaning there is a potential role for these genes in the necrosis. Indeed, marker e416, whose alleles were closely linked to the necrosis phenotype, has a gene encoding a disease resistance protein located within ~100kb. Although the data cannot show whether the Cerig DM genes involved in the necrosis are *R* genes, the fact that these QTLs contain *R* genes is an interesting point for future study. However, as QTLs map genetic traits that might consist of many linked genes within ~30cM of each other (Flint and Mott, 2001) it is possible that the QTLs identified by e217 and e416 may contain several potential DM-genes. For instance, in addition to the two *R* genes present in the region of the marker for the chromosome 9 QTL there is another interesting gene which encodes a UDP-glucosyltransferase that may act to regulate the expression of salicylic acid which has a key role in signalling pathogen-induced disease resistance (Alvarez, 2000).

As only markers for *R* gene clusters were screened, the choice of markers to screen for segregation with the hybrid necrosis phenotype to identify QTLs does not cover all of the

genome (four chromosomes did not have markers screened) and also potentially prevents the identification of genes other than *R* genes that might be involved. Although the genes involved in hybrid necrosis that have been functionally identified have all been involved in aspects of the immune response (Alcazar et al., 2009, Bomblies et al., 2007, Jeuken et al., 2009, Kruger et al., 2002, Masuda et al, 2007, Wulff et al., 2004, Yamamoto et al., 2010), it remains possible that genes of other functions could be involved. An improvement to this method would be to include other non-*R* gene markers and screen for markers on all chromosomes, not only those covered by the *R* gene markers that were selected.

The degree of hybrid necrosis in the segregating F_2 population of Cerig plants had a continuous distribution ranging from 0% to 75% with no distinct clustering. Despite this wide distribution in the degree of hybrid necrosis, the majority (95%, or 135 out of 142 individuals) of the F_2s had hybrid necrosis of 50% or lower. This skew in the distribution of necrosis, presumably due to the absence in the F_2 population of the incompatible alleles required to cause the highest levels of necrosis, can be attributed to the genotype of the F_1 individual that was randomly selected and selfed to generate the F_2.

One factor that could potentially affect the degree of hybrid necrosis due to the genotype of the F_1 parent is inbreeding depression as, given that this phenomenon lowers the fitness of individuals and reduces their ability to survive or reproduce, it could affect results arising from crossing experiments using an F_2 population generated by inbreeding. Clearly, if the F_1 individual that was selfed carried any deleterious alleles in heterozygous form then upon selfing these will become homozygous in some of the F_2. Indeed, crosses between Cop and some F_2s did produce a low seed set which could be attributed to inbreeding depression. However, inbreeding depression causing the loss of certain F_2 from the crossing population would only affect the degree of hybrid necrosis if the genes involved were linked to those causing hybrid necrosis.

Another explanation for how the genotype of the F_1 parent may affect the degree of hybrid necrosis caused by the F_2s is simply that the random selection of an F_1 individual from the Cer10 x Cer35 cross meant that the genome of this plant would have contained a random mixture of alleles at the DM-genes and so may not have included incompatible alleles at some of the loci and therefore neither does the F_2 population. Interestingly, the distribution appears to increase more rapidly after the ~40% rate which could be due to the cumulative effect of incompatible alleles or the presence of a less frequent allele which causes greater necrosis or because there are fewer individuals at this end of the distribution.

The wide, continuous range in the degree of hybrid necrosis caused by the F_2s is consistent with other examples of hybrid dysfunctions in plants, particularly necrosis, in which the variation

in phenotype severity is attributed to an allele dosage effect where the hybrid incompatibility phenotype increases in proportion to the number of incompatible alleles present (Bomblies et al., 2007, Hermsen, 1963, Ichitani et al., 2007, Kubo and Yoshimura, 2005, Song et al., 2009). As the necrosis increased when the necrosis alleles were present on both QTLs, this suggests an allele dosage effect in action.

Although a gene dosage effect could explain the wide variation in necrosis in the F_2 population as a whole, it cannot explain how the F_2s heterozygous for marker e416 (*Bb*) can give the same rate of necrosis as those homozygous for *B*. When crossed with Cop, only half the progeny inherit the incompatible *B* allele (assuming normal Mendelian inheritance) and so the expected necrosis rate for the *Bb* genotype is half that of *BB* i.e. half way between the necrosis rates of the two homozygous marker genotypes (~18%). The average necrosis of the heterozygotes at e416 is clearly higher than this (~26%).

There are several possible explanations for the equal rate of necrosis seen in the *BB* homozygote and the heterozygote. The first is that *B* is a marker for multiple alleles of a necrosis locus each of which have different degree of expression of the incompatibility factor. Such variation in the expressivity of alleles within a DM-gene, termed allelic expressivity, can cause variation in hybrid necrosis, as found in wheat (Hermsen, 1963). In this case, the incompatible allele *B* would in fact have multiple variants, each with a differing rate of expression of hybrid necrosis. The nature of the distribution of hybrid necrosis rates with the genotypes of these markers is similar to the continuous distribution of necrosis in wheat and is caused by allelic expressivity of two incompatible loci (Hermsen, 1963) giving some credence to this theory. Interestingly, *R* genes have multiple alleles (Michelmore and Meyers, 1998) which, if they are amongst the DM-genes in the Cerig population could be another factor explaining the wide continuous distribution seen in the segregating Cerig F_2 population.

Another explanation for the similar necrosis rate with the *BB* and *Bb* genotypes is that it is caused by segregation distortion (SD) which, in the cross between the heterozygous F_2 and Cop, could cause the *B* allele to be inherited from the F_2 parent in more than half of the progeny and therefore cause a higher than expected rate of hybrid necrosis from the heterozygous F_2. Although SD is in action, the observed ratio of genotypes suggests that SD distorts in favour of the *b* allele so it would be expected that out of the two alleles, this one would be present in more of the heterozygote F_2 x Cop progeny thus causing a lower than expected rate of necrosis. It would seem that SD, though clearly present, cannot explain the similar necrosis rates. It is not possible to say whether the distortion is a direct effect of the hybrid necrosis loci or whether it is a linked gene causing the unequal division of the alleles in the parental gametes. However, the presence of SD raises the possibility that the hybrid necrosis loci in the Cerig population could have arisen as a by-product of genetic conflict rather than a side effect of adaptive evolution.

There are other examples in the literature where genetic conflict has resulted in hybrid dysfunction (Johnson, 2010, Presgraves, 2010) although thus far the incompatibility phenotype has been hybrid sterility rather than lethality.

Alternatively, the similar necrosis rates could have arisen due to a sporophytic effect as the F_2s were always the female parent when crossed with Cop. Sporophytic effects can include genomic imprinting, where genes are silenced or expressed according to the DNA methylation pattern inherited from the maternal line, and cytonuclear incompatibility, a special type of DM-incompatibility between uniparentally inherited cytoplasmic genes and biparentally inherited nuclear genes (Barr and Fishman, 2010). However, during the characterisation of the hybrid necrosis phenotype in Chapter 3, reciprocal crosses were performed between Cerig and Cop to establish the presence of any such parent-of-origin effect and found no difference in the degree of necrosis which suggests this hybrid necrosis in this cross is not caused by aberrant nuclear-cytoplasmic interactions or imprinting.

Perhaps the most likely cause is that the average hybrid necrosis rates are misleading due to wide variation in necrosis within each genotype and differences in the number of F_2 with each genotype. The wide distribution of necrosis (from 0% to 53.33%) with the small sample size of F_2s with the *BB* genotype (eleven individuals) is likely to be an important factor in terms of establishing an average hybrid necrosis rate for this genotype and could be misleading and could conceal differences between the *BB* and *Bb* genotypes. For instance, there were four low-necrosis rate F_2s (0%, 12.17%, 13.33% and 16.87%) with a *BB* genotype. These rates of necrosis, although surprisingly low, appear to be accurate as they did not have a significant difference between their repeats, although the one with 0% necrosis only had one repeat.

It is surprising that the *BB* genotype is associated with some very low rates of necrosis locus. Interestingly, Hermsen (1963) also noted that in wheat hybrid necrosis some F_1s remained phenotypically normal throughout their development in spite of their genotype of necrosis. It is known that there are a small number of genes in Cerig involved in hybrid necrosis (Macnair & Christie, 1983) and that the effects of genic incompatibilities are often additive (Coyne and Orr, 2004). As such, the gene on the chromosome 9 QTL may occur with another gene that affects necrosis, perhaps a gene of larger effect, so although an individual has the necrosis allele at this locus it has a low necrosis rate due to the presence of a non-necrotic allele at another locus.

The genotype for the homozygote necrosis allele of marker e279, *782* (*C*) did not appear. The *C* allele only existed in the heterozygous form, the most likely reason being that the *C* allele is linked to a recessive lethal gene and so no viable individuals possess this genotype. Lethal alleles are the most common class of mutation, perhaps reflecting the fundamental functions of

genes as controlling the processes essential to the growth and development of organisms, and have been found in a multitude of organisms including *Drosophila* (Mestres et al., 2009), *C. elegans* (Zipperlen et al., 2001), humans, causing diseases such as cystic fibrosis and sickle cell anaemia, mice (Dunn, 1956) and flowering plants (Macnair, 1993), and bring about the death of the organism carrying it. Moreover, in most cases, in higher diploids the lethal alleles are usually recessive and so are expressed only in homozygotes whilst dominant lethal alleles, expressed in heterozygotes, are rapidly eliminated and thus rarely detected.

Although 200 F_2 plants were grown, only half (109 individuals) were successfully scored for necrosis and genotypes for both markers due to pest problems and mortality reducing the number of F_2s available for crosses to Cop, mildew infections causing difficulties scoring for hybrid necrosis, and thirty-three individuals in which PCR reactions failed to amplify a fragment with marker e279. Of the five combined genotypes for e416 and e279 represented amongst these 109 F_2s, some contained a very small number of individuals making the accuracy of their average necrosis rate questionable. Clearly, a larger overall sample size, such as the originally intended 200 F_2s would likely yield far more informative and definite results. It was possible that the statistical significance threshold used to select the 142 F_2 individuals used in the marker screening (see 6.3.1) was too conservative. However, performing a Bonferroni correction for multiple comparisons on the chi-squared results for the “all trays” column in Appendix 6.1 reveals only an additional six individuals that could have been used in the marker screening, indicating that the baseline used was reasonable but slightly too conservative.

The error in the initial analysis of the hybrid necrosis data meant that seven plants were incorrectly used in the BSA which, because they were not at the extremes of the necrosis phenotype, could have resulted in failure to reveal markers that co-segregate with hybrid necrosis. However, having analysed the BSA data both with and without the incorrect plants, the significance of only one marker, e279, was affected and was found to be significant when all the BSA plants were analysed but not significant when the incorrect plants were omitted. The one-way ANOVA performed on all plants genotyped at this marker was significant suggesting that the presence of these incorrect plants has not undermined the strength of BSA. By modifying the true BSA methodology and keeping the DNA samples of individuals at the extreme phenotypes separate, the impact of errors such as this was limited.

The use of the incorrect plants in the BSA arose as a result of two factors: miscalculations when working out the average necrosis rate and significant differences between the necrosis rates. These significant differences were caused mainly by two factors both described during the characterisation work in Chapter 3 and apply to the differences found between all repeats. First, the occurrence of a powdery mildew infection on the hybrids plants which caused a yellowing and necrosis on the leaves which was very similar to the hybrid necrosis phenotype

made scoring accurately for hybrid necrosis difficult. As far as possible measures were taken to limit the occurrence of mildew, but time and space considerations were restrictive factors.

Second, it seems likely that the hybrid necrosis phenotype is alleviated by higher temperatures which could explain some of the large differences in the degree of necrosis found between replicate crosses sown at different times of the year. The differences in hybrid necrosis between replicate crosses for each F_2 showed a general trend. Those which had their first replicate sown in March/April tended to have higher necrosis than their replicates sown in June/July whilst those whose first replicate was sown in July/August tended to have lower necrosis than their replicates sown in September. This is consistent with necrosis being temperature sensitive. However, significant differences in necrosis rates also seem to appear in clusters according to dates which could correspond to particularly bad mildew infections. The impact of this temperature sensitivity is important. For example, two plants were incorrectly used in the BSA ($F_2$3 and $F_2$6) due to necrosis rates in their replicate crosses being highly different which was initially attributed to mis-scoring due to mildew damage and so the lower necrosis rates of the second replicate was used (0%) but it was likely that the warmer ambient temperature caused an alleviation of the necrosis phenotype. To improve the accuracy of measuring the rate of hybrid necrosis, a tighter control on the temperature of the growing environment is required. Interestingly, however, a temperature-sensitive effect is consistent with the possible role of disease-resistance genes, known to be present on the QTLs identified in this research, as the plant immune system is suppressed by higher temperatures (Bomblies and Weigel, 2007) and presents an exciting direction for future research.

6.5 Conclusion

Reproductive isolation can arise from the additive effects of multiple genic incompatibilities and a hybrid might need to have particular alleles at three or more loci before its fertility or viability is affected (Coyne and Orr, 1998, 2004, Orr, 1995) and so, due to the polymorphism within the hybrid incompatibility genes in the Cerig population, the strength of the barrier (i.e. the degree of hybrid necrosis) between Cerig and Cop depends on the specific combination of alleles at all the loci involved in the DM incompatibility interaction.

This chapter has made the preliminary steps towards elucidating these incompatibility genes in Cerig and has identified two QTL, one on chromosome 9 and an enhancer on chromosome 12, that act additively to cause ~30% of the hybrid necrosis. Interestingly, both of these QTLs contain at least one disease resistance gene and the hybrid necrosis phenotype appears to be temperature sensitive, implying a possible role for the plant immune system in this crossing

barrier. Although there are obviously additional loci involved, this is the first step towards locating the genomic position of the genes involved in the Cerig population. Ultimately, identifying all the loci that contribute to the Cerig side of the incompatibility with Cop would enable a greater understanding of how complex interactions between several incompatible genes can bring about hybrid inviability.

Chapter 7

General Discussion

This thesis examines the genetic basis of an intrinsic postzygotic reproductive isolating barrier brought about by deleterious epistatic interactions between incompatible alleles of complementary genes (Dobzhansky-Muller incompatibilities) in crosses between two populations of *Mimulus guttatus*: copper-tolerant 'Copperopolis' (Cop) and copper non-tolerant 'Cerig'. As such, this thesis builds upon the findings of Macnair & Christie (1983) which demonstrated three key points about this crossing barrier: first, that in Cop a single locus is involved in the incompatibility and that this is either tightly linked or synonymous with the locus conferring copper tolerance; second, that the Cerig population is polymorphic for the genes involved; and third, that there are a small number of genes involved in Cerig. Described by Macnair & Christie (1983) as a synthetic lethal, the phenotype of this hybrid incompatibility (see Chapter 3) is consistent with the characteristics of 'hybrid necrosis', which include tissue yellowing and necrosis, dwarfism, and, in severe cases, lethality.

7.1 The single hybrid necrosis locus in Copperopolis

The copper tolerance and hybrid necrosis phenotypes were established as being controlled by discrete loci, *T* and *NEC* (Chapter 4). Due to their tight linkage, until now it has not been known whether these loci were distinct or synonymous but the identification of two plants in which recombination had disrupted the linkage has resolved this issue. Although these plants gave conclusive results and clear evidence of the discrete nature of the loci, the result would have been strengthened by identifying more individuals of a similar genotype and phenotype combination, although the tight linkage between the two loci does make it unlikely for many recombinants to occur. However, out of a potential sixty-eight plants, only twenty-seven were tested due to time constraints, plant mortality, pest problems, phenotype ambiguity and delayed marker genotyping. Thus, including these forty-one additional plants may have improved the chances of finding additional cases where recombination had disrupted the linkage between the phenotypes.

The discrete nature of the *NEC* and *T* loci has interesting implications in terms of the evolution of this crossing barrier and how this hybrid necrosis gene has arisen and become fixed in the Cop population. Macnair & Christie (1983) suggested that hybrid necrosis could be a pleiotropic

effect of the tolerance locus and would therefore be a by-product of direct selection on an adaptive trait. However, now that it is known that these are discrete loci this can be eliminated as a possibility.

There are two possible alternative mechanisms by which *NEC* became fixed. First, if *NEC* confers a selective advantage it is possible that it has been directly selected for. In nature, genes causing sterile or lethal hybrids between diverging populations are not likely to be directly selected for as natural selection is unlikely to select for maladaptive traits (Darwin's 'valley of low fitness'). However, as the Cerig and Cop populations are geographically isolated, hybridisation between them would not occur in nature and so the obstacles to direct selection posed by a maladaptive hybrid are extraneous. As such, selection could act directly upon *NEC* in this system. Then again, the Cerig genes were also found in another population, Napoleon, which is only 10km from the Copperopolis population (Christie and Macnair, 1987) meaning that the interacting genes may be in the same geographical area.

The second explanation is that due to its very tight linkage to the tolerance (*T*) locus, *NEC* was carried to fixation through hitchhiking when selection acted on the *T* locus. Interestingly, current research at Duke, which is at an early stage and has not yet been substantiated, shows signs of a selective sweep on the tolerance locus thus supporting the hitchhiking theory for *NEC* fixation (Kevin Wright, personal communication). More marker screening is needed to determine whether this is so, but early indications certainly suggest that Macnair & Christie's (1983) suggestion that hybrid necrosis arose as a by-product of selection for the tolerance locus, albeit as hitchhiking rather than pleiotrophy, is correct.

Having established that hybrid necrosis is caused by a discrete locus, fine-mapping enabled the identification of the region containing the *NEC* locus as a 59kb region on chromosome 9 of the *M.guttatus* genome. Bioinformatic analysis of this region revealed the presence of two definite genes encoding proteins with putative conserved domains, one a DUF563 glycosyltransferase, the other a Jumonji-domain protein, both of which have potential as candidates for *NEC* as discussed in detail in Chapter 5. A further open reading frame was identified as encoding a domain belonging to the exonuclease/ endonuclease/ phosphatase superfamily, although there is some uncertainty surrounding this as the mRNA sequence identified during the genome sequencing does not appear to have conserved domains. While a role in hybrid necrosis is plausible for all of these genes, particularly the glycosyltransferase, some interesting preliminary results emerging from the continued fine-mapping of this region by Duke University indicates that the region containing *NEC* is 283-297kb (Kevin Wright, personal communication), therefore excluding the JmjC protein and implicating the glycosyltransferase as *NEC*.

Although the target protein(s) of this enzyme is not known, the fact that glycosylation is recognized as performing a critical role in plant disease resistance, both in pathogen receptor protein conformation (Haweker et al., 2010) and salicylic acid regulation (Alvarez, 2000), is interesting given that there are several lines of support for the involvement of the plant immune system in this crossing barrier. Firstly, a very similar phenotype was found between hybrids suffering hybrid necrosis and plants suffering from mildew infection, which would imply an active immune response. Secondly, the phenotype was strongly similar to that described in the literature where genes involved in the disease response have been conclusively shown to be involved in the hybrid necrosis (Alcazar et al., 2009, Bomblies et al., 2007, Jeuken et al., 2009, Kruger et al., 2002, Masuda, 2007, Wulff et al., 2004, Yamamoto et al., 2010). Thirdly, although not conclusively demonstrated, the hybrid necrosis phenotype appeared to be alleviated at higher temperatures, which is a well documented feature of hybrid necrosis resulting from aberrant immune system activity (Zhu et al., 2010 and the preceeding references).

As such, it is feasible that the lack of functional ability of this glycosyltransferase in the hybrid genome causes aberrant glycosylation activity that activates the disease response in the absence of a pathogen thereby causing the hybrid necrosis phenotype. There are a number of ways in which *NEC*, as a glycosyltransferase, might cause immune-like necrosis in hybrids. For instance, there is evidence that plant pattern recognition receptors that recognise pathogens are subject to glycosylation that is essential for their function in triggering plant immunity (Haweker et al., 2010). However, this implies that a non-functional glycosyltransferase would diminish the immune response so it is hard to reconcile how this activity could cause aberrant activation of the immune system. A more conceivable possibility comes from the fact that glycosylation regulates salicylic acid (SA) which has a key role in the activation of the disease resistance that is frequently associated with the hypersensitive response (HR) characterised by localised programmed cell death (PCD) (Dangl et al., 1996) and enhances defence responses promoting lesion formation and induces systemic acquired resistance (SAR) (Alvarez, 2000). It is thought that glucosylation (transfer of glucose residues usually from UDP) represents detoxification of SA (Lim et al., 2002). Therefore, if the NEC encodes a glycosyltransferase that regulates SA activity but was non-functional in the hybrid genome, the disease resistance response could be activated. However, how this mechanism may involve the NEC protein interacting with the products of a small number of Cerig genes is a little perplexing.

It is possible that NEC is able to perform an essential glycosylation activity in Cop but not in the hybrid genome, perhaps because divergence has caused altered glycosylation site specificity between NEC and the Cerig proteins. Understanding how NEC may interact with the proteins encoded by the small number of genes in Cerig to cause hybrid necrosis inevitably depends on identifying the function of these complementary genes in Cerig.

7.2 Identifying hybrid necrosis QTLs in Cerig

This thesis sought to identify the QTLs underpinning the incompatible alleles contributed by Cerig to this crossing barrier. The degree of necrosis (i.e. the number of hybrids that were affected) is dependent on the individual Cerig parent used in the cross with Cop, thus supporting Macnair & Christie's (1983) conclusion that the Cerig population is polymorphic so that the genotypes *AA*, *Aa* and *aa* all exist for a given gene and that there are a small number of genes involved. The polymorphism at multiple loci led to a wide distribution in the degree of hybrid necrosis and also results in this barrier only partially restricting gene flow between Cop and Cerig.

The way in which the loci in Cerig interact with *NEC* is important in terms of how the Cerig polymorphism will affect the degree of necrosis. If each Cerig locus interacts independently with *NEC* in a several two-locus interactions, then a Cerig plant carrying a single incompatible allele at one locus could cause some degree of necrosis. However, if the interaction involves three or more loci, then this same Cerig individual might not cause any necrosis. Furthermore, there could be an allele dosage effect so that the more incompatible alleles there are, the greater the degree of necrosis.

This thesis has identified two QTLs that contribute to the incompatibility arising from the Cerig population. One QTL maps to chromosome 9 and causes around 20% of the hybrid necrosis found in this cross. The second QTL maps to chromosome 12 and acts as an enhancer of the QTL9 incompatibility, affecting the degree of necrosis only when the incompatible allele is present on QTL9 whereupon it causes an additional 10% of necrosis.

As such, assuming QTL9 harbours a single incompatibility gene and that there was not another locus involved on an unidentified QTL, the data indicates that incompatibility occurs between *NEC* and the locus on chromosome 9 and does not require the involvement of a third locus for necrosis to be expressed. This could therefore represent a simple, two gene interaction as frequently found in other systems (Lowry et al, 2007). However, the incompatibility from QTL12 is only expressed when the incompatibility is present from QTL9 implying that hybrid necrosis in this cross is not simply an accumulation of multiple two-gene interactions between *NEC* and the Cerig loci.

How representative this is of the other Cerig loci is impossible to say. It could be that the other loci each interact independently with *NEC* and their effects are additive which then accumulate to cause the more severe forms of necrosis, or they could interact in a more complex manner.

As *NEC* is not a plant immune gene, it is exciting to find that both the QTLs identified harbour disease resistance (*R*) genes, thereby providing the first indication that this crossing barrier might be the result of aberrant activation of the plant immune system. Hybrid necrosis due to incompatibilities involving *R* genes has been found in other systems (Bomblies et al, 2007; Alcazar et al, 2009; Kruger et al, 2002; Yamamoto et al, 2010) and as *R* genes tend to evolve quickly due to strong selection arising from pathogen evolution they are good candidates for DM genes.

Furthermore, there are two mechanisms by which R proteins become activated that may involve glycosylation or phosphorylation, thus implicating the DUF563 glycosyltransferase and the mRNA that may encode a phosphatase. First, the guard hypothesis proposes that R proteins are activated upon recognition of a complex formed when a pathogen effector interacts with a guardee protein (van der Biezen and Jones, 1998). This has been found in *Arabidopsis* where the RIN4 guardee protein is posttranslationally modified by the pathogen effector which in turn activates the R proteins (Belkhadir et al., 2004). A homolog of RIN4 was found to be involved in crosses between lettuce species that show hybrid necrosis (Jeuken et al., 2009). Although the posttranslational modification of RIN4 was hyperphosphorylation (Belkhadir et al., 2004), which could implicate the mRNA at 282kb as *NEC*, glycosylation is also a well known posttranslational modification. As such, in the hybrid genome the NEC protein encoded by Cop could aberrantly glycosylate guardee proteins causing the activation of guard R proteins. Two other R proteins belonging to a class thought to function as guards have been implicated in hybrid necrosis (Dangl and Jones, 2001, Kruger et al., 2002, Wulff et al., 2004). The second mechanism for R protein activation involves pathogen effectors mimicking plant transcription factors and targeting specific promoters which turn on plant host genes and, in resistant plants, *R* genes. The activity of transcription factors is regulated by posttranslational modifications, such as glycosylation or phosphorylation (Meshi and Iwabuchi, 1995). It is possible then, that in the hybrid genome NEC aberrantly modifies a transcription factor which then binds to promoters of R genes causing their expression in the absence of a pathogen.

Although the presence of *R* genes on the QTLs is highly interesting and several mechanisms for them as the complementary genes for NEC are plausible, without the definite identity of any of the genes, developing a model for how the incompatible interactions occur and cause hybrid necrosis is somewhat ambitious.

Some aspects of the data from the mapping population of Cerig were curious. Firstly, no homozygotes for the incompatible allele were found on QTL12 which was considered most likely to be due to a recessive lethal allele linked to the locus for incompatibility. Secondly, the similarity in the degree of hybrid necrosis caused by F_2 parents that were either heterozygous or homozygous for the incompatibility allele on QTL9 was unexpected given that all offspring from

the latter should carry the incompatible allele and therefore express necrosis whilst only half the progeny from the former should be necrotic. Several explanations for this were considered. First, there could be multiple allelic variants of the incompatibility locus on QTL9, each of which confers a different degree of expression of the incompatibility factor that causes hybrid necrosis. Such allelic expressivity can cause variation in hybrid necrosis, as found in wheat (Hermsen, 1963) which shows a continuous distribution of necrosis similar to that found in this cross giving some credence to this theory. Furthermore, *R* genes are known to have multiple alleles (Michelmore and Meyers, 1998), an interesting fact given the presence on QTL9 of an *R* gene and the strong potential for *R* gene involvement in hybrid necrosis. Second, as the F_2s were always the female parent when crossed with Cop a sporophytic effect might be in action which could cause altered genomic imprinting or cytonuclear incompatibility in the hybrid genome (Barr and Fishman, 2010). However, the characterisation of the hybrid necrosis phenotype in Chapter 3 established there was no effect of the maternal parent. Third, as Cerig individuals possess a small number of loci that contribute to hybrid necrosis and only two were identified in this thesis, it is possible that another necrosis locus which was not identified was present in the F_2 genome. If the effect of this locus was unlinked to the genotype at QTL9, this could cause the degree of necrosis associated with the three genotypes to be distorted. Fourth, and perhaps most likely, the similar necrosis rates could simply be due to the wide range in the degree of necrosis and a small number of homozygotes for the incompatible allele resulted in a misleadingly low average necrosis rate. Indeed, the distribution of the hybrid necrosis rates with the three genotypes of the QTL9 marker (Chapter 6, figure 6.7) indicates that the necrosis rates of the heterozygote do lie in between the two homozygotes, but that there are significantly more heterozygotes than incompatible homozygotes.

There also appeared to be evidence of segregation distortion (SD) at QTL9 resulting in over-representation of the compatible allele. Whether or not this SD is due to the same locus that causes hybrid necrosis is unknown; it may be that the marker for hybrid necrosis associated with this QTL is also linked to a separate locus that causes SD. However, the occurrence of SD is interesting given that it has been reported in several other cases of hybrid dysfunctions (reviewed by Johnson, 2010, Presgraves, 2010). SD is caused by selfish genetic elements, termed segregation distorters, that manipulate host reproduction to facilitate their own transmission to the next generation at the expense of other genes, or even the organism. To suppress the selfish DNAs or compensate for their deleterious effects other host genes often evolve (Burt and Trivers, 2006, Presgraves, 2010). This gives rise to genomic conflict, which can lead to the evolution of hybrid dysfunctions either due to incompatibilities between host genes that have evolved to suppress the effects of the selfish genetic elements (SGEs) or because SGEs are not suppressed in the genome of the hybrid (Burt and Trivers, 2006, McDermott and Norr, 2010, Presgraves, 2010). This latter mechanism, of which there are three examples in *Drosophila* (see Presgraves, 2010 for a review), is particularly interesting as it

proposes an alternative mechanism to the current classical view of hybrid dysfunction arising from genes that have diverged and whose new alleles are incompatible. Genomic conflict is considered to have caused the evolution of several cases of hybrid sterility (Barr and Fishman, 2010, Case and Willis, 2008, Oliver et al., 2009, Phadnis and Orr, 2009, Presgraves, 2010) but has not yet been identified in any hybrid lethality/necrosis systems. However, if genetic conflict is the force driving the evolution of hybrid dysfunctions, then the first steps in the evolution of hybrid sterility or inviability may be nearly neutral or even deleterious rather than adaptive (Presgraves, 2010).

Although two QTLs were identified, overall the mapping of hybrid necrosis QTLs in Cerig was disappointing in that the two QTLs that were identified can only explain around 30% of the necrosis and the total number of QTLs, and therefore genes, is still unknown. This was caused by two factors that are not mutually exclusive.

First, the F_1 individual randomly selected to generate the F_2 mapping population did not, by chance, have incompatible alleles present at all the Cerig genes. This could have been resolved by testing the necrosis rates of the F_1s and selecting the individual with the highest rate. However, as determining the degree of necrosis requires performing crosses and counting the number of offspring with necrosis symptoms, time limitations prevented this from being practical.

Second, the strategy to target marker screening at *R* gene clusters was too limited in two ways. Firstly, it did not cover the genome adequately. The markers tested (as shown in Appendix 7.1) covered around 525cM of the ~1450cM *Mimulus* genome (estimated using the *Mimulus* genome map and by assuming everything within 20cM of the markers was covered). This is equivalent to only 36% of the genome. The markers were located on ten of the fourteen chromosomes meaning ~380cM or 27% of the genome was omitted from the research. Secondly, it potentially prevented the detection of loci other than *R* genes that contributed to the hybrid necrosis phenotype. Adopting a traditional random marker screening approach would have provided a more inclusive and adequate coverage of the genome and would in all probability have enabled the identification of additional markers and QTLs.

There are, however, many interesting directions in which future work could progress.

7.3 Future Work

For the Cop side of the crossing barrier, firmly establishing the identity of *NEC* is essential. Although fine-mapping has restricted the number of candidate genes, perhaps the best way to achieve this would be to knock out this gene and then perform crosses to a plant known to normally result in high levels of hybrid necrosis and see whether its absence alleviates the necrotic phenotype. Alternatively, transformations could be performed in which a plant that is known to cause no hybrid necrosis is transformed with the *NEC* gene and then re-tested for its propensity to cause necrosis. However, genome manipulation of *Mimulus* is not easy as the standard methods do not work well. An alternative direction would be to perform expression analysis to see what genes are up-/down-regulated in necrotic and non-necrotic hybrids and are therefore involved in the process of necrosis. This would give indications as to the physiological factors and may reinforce the functional identities of genes implicated.

In the Cerig population, the work in this thesis could be expanded upon by identifying the genes on the two QTLs found. This could be achieved by developing markers for fine-mapping. Another important direction is to identify the QTLs that account for the remaining ~70% of the necrosis by improving the coverage of the genome searched for QTLs by using random markers and screening markers not targeted at *R* gene clusters.

Once all the loci that contribute to the Cerig side of the incompatibility with Cop have been identified then it will be possible to begin to build a model of the genetic architecture and how it brings about hybrid necrosis.

In addition to identifying the complementary genes from the two populations, there are other interesting avenues to pursue, particularly the temperature sensitivity of the necrosis phenotype and factors causing the variable severity and time of onset of the first necrosis symptoms.

It is probable that variable greenhouse temperatures during the hybrid growing experiments affected the rate of hybrid necrosis and thus caused the data discrepancies reported in this thesis. Thus, to facilitate further study and improve the accuracy of measuring the rate of hybrid necrosis, it would be prudent to establish whether hybrid necrosis in this cross is temperature sensitive, and if so what the threshold temperature is for alleviating the phenotype. Furthermore, this would be an interesting aspect of the phenotype with regards to the potential involvement of the immune system. It would also be interesting to see whether there is a time limit for the length of exposure to lower temperatures a hybrid can endure before the higher temperature is no longer able to alleviate necrosis, as has been found in *Nicotiana* (Masuda et al, 2007) or whether the various severities of the phenotype are affected differently.

7.4 Final Conclusion

The acquisition of reproductive isolating barriers is essential in the divergence and maintenance of species and there is growing research interest in the molecular basis of this phenomenon of intrinsic postzygotic isolation brought about by Dobzhansky-Muller incompatibilities. One particular barrier of increasing interest is hybrid necrosis in plants. Not only is this interesting in terms of evolution but also in terms of plant breeding, for example in crop plants, such as wheat, hybrid necrosis can pose a serious problem in that it prevents the combining of desirable traits from different genotypes or transferring gene from related species to commercial cultivars (Chu et al., 2006).

Yet, in spite of this interest still relatively few examples of genes involved in these incompatibilities have been identified and those that have been tend to be involved in two-gene interactions. As such, relatively little is known about the allelic effects resulting from more complex interactions between three or more loci, as is potentially the case with the cross between Cop and Cerig.

By demonstrating that copper tolerance and hybrid necrosis are controlled by discrete loci, that the gene in Copperopolis that causes necrosis is either a glycosyltransferase, a Jumonji-domain protein, or possibly a phosphatase, with recent data suggesting the former, that two of the loci in Cerig are on QTLs containing *R* genes and that one acts as an enhancer of the other to cause approximately 30% of the necrosis, this thesis has contributed to furthering the understanding of the genetic basis of this crossing barrier which is an example of a Dobzhansky-Muller incompatibility involving multiple loci.

Ultimately, regardless of the mechanism by which hybrid necrosis is brought about in this cross, the polymorphism in the Cerig population that causes necrosis to range from almost 100%, so that gene flow is almost entirely eliminated and consequently RI is almost total, to 0% so that genes can flow freely between the populations and so RI is almost non-existent, means this intrinsic postzygotic isolating barrier can only partially prevent gene flow between these two populations. As such, this may represent an early-stage reproductive isolation barrier in the evolution of RI between these two *M. guttatus* populations. As these populations continue to diverge they may accumulate more incompatible alleles that will eventually lead to complete postzygotic reproductive isolation.

Appendix 4.1

Table A4.1 Raw data for hybrid necrosis from crosses between Cerig 10 or Cerig 35 and the tolerant (TR) and non-tolerant (NTR) recombinants

Cross	No. of Healthy hybrids	No. of Necrotic hybrids	Total no. of hybrids scored	% Necrosis
Cer 10 & TR				
25_A09 x Cer 10 (10)	26	34	60	56.67
25_A09 x Cer 10 (11)	12	48	60	80.00
Cer 10 (11) x 25 _A09 1	29	31	60	51.67
Cer 10 (11) x 25 _A09 2	25	35	60	58.33
26_C07 x Cer 10 (1)	34	26	60	43.33
26_C07 x Cer 10 (1/2/4)	29	31	60	51.67
26_C07 x Cer 10 (2)	27	33	60	55.00
26_C07 x Cer 10 (6)	39	21	60	35.00
Cer 10 (1) x 26_C07	34	26	60	43.33
Cer 10 (2) x 26 _C07	37	23	60	38.33
Cer 10 (2) x 26_C07	42	18	60	30.00
Cer 10 (6) x 26_C07	30	30	60	50.00
26_H02 x Cer 10 (7)	26	34	60	56.67
26_H02 x Cer 10 (7)	22	38	60	63.33
26_H02 x Cer 10 (8)	28	32	60	53.33
26_H02 x Cer 10 (8)	28	32	60	53.33
Cer 10 (7) x 26_H02	23	37	60	61.67
Cer 10 (7) x 26_H02	28	32	60	53.33
Cer 10 (8) x 26_H02	44	16	60	26.67
Cer 10 (8) x 26_H02	39	21	60	35.00
28_G06 x Cer 10 (1)	36	24	60	40.00
28_G06 x Cer 10 (2)	37	23	60	38.33
28_G06 x Cer 10 (6)	27	33	60	55.00
Cer 10 (1) x 28_G06	33	27	60	45.00
32_G02 x Cer 10 (2)	24	36	60	60.00
32_G02 x Cer 10 (4)	18	42	60	70.00
32_G02 x Cer 10 (5)	25	35	60	58.33
32_G02 x Cer 10 (6)	26	34	60	56.67
Cer 10 (1) x 32_G02	21	39	60	65.00
Cer 10 (2) x 32_G02	19	41	60	68.33
Cer 10 (5) x 32_G02	25	35	60	58.33
Cer 10 (6) x 32_G02	29	31	60	51.67
34_A03 X Cer 10 (1)	30	30	60	50.00
34_A03 x Cer 10 (2)	29	31	60	51.67
34_A03 x Cer 10 (5)	10	20	30	66.67
34_A03 x Cer 10 (6)	24	36	60	60.00
Cer 10 (1) x 34_A03	25	35	60	58.33
Cer 10 (1) x 34_A04	34	26	60	43.33
Cer 10 (2) x 34_A03	28	32	60	53.33
Cer 10 (6) x 34_A03	39	21	60	35.00
35_F09 x Cer 10 (1)	33	27	60	45.00
35_F09 x Cer 10 (2)	8	22	30	73.33
Cer 10 (1) x 35_F09	29	31	60	51.67

Cross	No. of Healthy hybrids	No. of Necrotic hybrids	Total no. of hybrids scored	% Necrosis
Cer 10 (1) x 35_F09	26	34	60	56.67
43_G01 x Cer 10 (10)	28	32	60	53.33
43_G01 x Cer 10 (11)	26	34	60	56.67
Cer 10 (10) x 43_G01	17	43	60	71.67
Cer 10 (11) x 43_G01	34	26	60	43.33
43_H01 x Cer 10 (10)	21	39	60	65.00
43_H01 x Cer 10 (11)	30	30	60	50.00
Cer 10 (11) x 43_H01 1	30	30	60	50.00
Cer 10 (11) x 43_H01 2	32	28	60	46.67
53_B01 x Cer 10 (10)	29	31	60	51.67
53_B01 x Cer 10 (11)	28	32	60	53.33
Cer 10 (11) x 53_B01 1	26	34	60	56.67
Cer 10 (11) x 53_B01 2	31	29	60	48.33
	1569	**1731**	**3300**	**52.45**
Cer 10 (8) x 25_E11	*56*	*4*	60	6.67
Cer 10 (7) x 25_E11	*52*	*8*	60	13.33
Cer 10 (8) x 25_E11	*24*	*6*	30	20.00
25_E11 x Cer 10 (7)	*54*	*6*	60	10.00
25_E11 x Cer 10 (7)	*53*	*7*	60	11.67
25_E11 x Cer 10 (8)	*44*	*16*	60	26.67
	283	**47**	**330**	**14.24**
TR x Cer35 (Control)				
26_C07 x Cer 35 (1)	51	9	60	15.00
28_G06 x Cer 35 (2)	46	14	60	23.33
32_G02 x Cer 35 (3)	59	1	60	1.67
34_A03 x Cer 35 (4)	55	5	60	8.33
35_F09 x Cer 35 (1)	49	11	60	18.33
	260	**40**	**300**	**13.33**
Cer 10 & NTR				
10_F08 x Cer 10 (10)	59	1	60	1.67
10_F08 x Cer 10 (11)	56	4	60	6.67
Cer 10 (10) x 10_F08	57	3	60	5.00
Cer 10 (12) x 10_F08	56	4	60	6.67
14_D05 x Cer 10 (10)	53	7	60	11.67
14_D05 x Cer 10 (11)	58	2	60	3.33
Cer 10 (10) x 14_D05	60	0	60	0.00
Cer 10 (11) x 14_D05	58	2	60	3.33
19_C03 x Cer 10 (7)	60	0	60	0.00
19_C03 x Cer 10 (9)	51	9	60	15.00
Cer 10 (7) x 19_C03	56	4	60	6.67
Cer 10 (8) x 19_C03	26	4	30	13.33
19_F06 x Cer 10 (7)	59	1	60	1.67
19_F06 x Cer 10 (8)	58	2	60	3.33
Cer 10 (7) x 19_F06	54	6	60	10.00
Cer 10 (7) x 19_F06	51	9	60	15.00
Cer 10 (8) x 19_F06	59	1	60	1.67
19_H12 x Cer 10 (7)	57	3	60	5.00
Cer 10 (7) x 19_H12	57	3	60	5.00
Cer 10 (8) x 19_H12	54	6	60	10.00
24_D09 x Cer 10 (7)	57	3	60	5.00
24_D09 x Cer 10 (8)	59	1	60	1.67
Cer 10 (7) x 24_D09	57	3	60	5.00

Appendix 4.1: Raw data for hybrid necrosis from crosses between Cerig and tolerant or non-tolerant recombinants

Cross	No. of Healthy hybrids	No. of Necrotic hybrids	Total no. of hybrids scored	% Necrosis
Cer 10 (7) x 24_D09	52	8	60	13.33
Cer 10 (8) x 24_D09	53	7	60	11.67
29_G11 x Cer 10 (10)	60	0	60	0.00
29_G11 x Cer 10 (11)	58	2	60	3.33
Cer 10 (11) x 29_G11 1	26	4	30	13.33
Cer 10 (11) x 29_G11 2	54	6	60	10.00
32_H09 x Cer 10 (2)	54	6	60	10.00
32_H09 x Cer 10 (4)	54	6	60	10.00
Cer 10 (1) x 32_H09	53	7	60	11.67
Cer 10 (1) x 32_H09	58	2	60	3.33
Cer 10 (2) x 32_H09	58	2	60	3.33
Cer 10 (6) x 32_H09	60	0	60	0.00
35_A12 x Cer 10 (1)	57	3	60	5.00
35_A12 x Cer 10 (2)	52	8	60	13.33
Cer 10 (1) x 35_A12	52	8	60	13.33
Cer 10 (1) x 35_A12	60	0	60	0.00
Cer 10 (2) x 35_A12	53	7	60	11.67
Cer 10 (5) x 35_A12	60	0	60	0.00
35_B06 x Cer 10 (7)	60	0	60	0.00
Cer 10 (7) x 35_B06	42	18	60	30.00
41_C05 x Cer 10 (10)	45	0	45	0.00
41_C05 X Cer 10 (11)	60	0	60	0.00
Cer 10 (10) x 41_C05	53	7	60	11.67
Cer 10 (10) x 41_C05	55	5	60	8.33
Cer 10 (11) x 41_C05	54	6	60	10.00
43_B01 x Cer 10 (10)	58	2	60	3.33
43_B01 x Cer 10 (11)	59	1	60	1.67
Cer 10 (10) x 43_B01	54	6	60	10.00
43_C01 x Cer 10 (10)	59	1	60	1.67
43_C01 x Cer 10 (11)	58	2	60	3.33
Cer 10 (11) x 43_C01	57	3	60	5.00
Cer 10 (11) x 43_C01	56	4	60	6.67
46_B12 x Cer 10 (10)	60	0	60	0.00
46_B12 x Cer 10 (11)	56	4	60	6.67
48_B01 x Cer 10 (7)	58	2	60	3.33
48_B01 x Cer 10 (8)	59	1	60	1.67
Cer 10 (7) x 48_B01	27	3	30	10.00
Cer 10 (8) x 48_B01	26	4	30	13.33
	3302	**223**	**3525**	**6.33**
Cer 10 (1) x 25_E01 1	41	19	60	31.67
Cer 10 (1) x 25_E01 2	38	22	60	36.67
Cer 10 (2) x 25_E01	40	20	60	33.33
Cer 10 (6) x 25_E01	36	24	60	40.00
25_E01 x Cer 10 (2)	21	39	60	65.00
25_E01 x Cer 10 (5)	36	24	60	40.00
	212	**148**	**360**	**41.11**
NTR x Cer35 (Control)				
25_E01 x Cer 35 (4)	59	1	60	1.67
32_H09 x Cer 35 (2)	59	1	60	1.67
35_A12 x Cer 35 (4)	55	5	60	8.33
	173	**7**	**180**	**3.89**

Appendix 6.1

Chi-squared values between replicates in the F_2 x Cop crosses

Table A6.1 Chi-squared values for testing for homogeneity between replicates trays from the same F_2 x Cop cross (columns 2 & 3), replicate F_2 x Cop crosses (column 4) and between all replicate trays each F_2 (column 6)

** = significant p < 0.05; ** = significant p < 0.01; † significant after Bonferroni correction for multiple comparisons*

F_2	χ^2 between Trays (cross 1)	χ^2 between Trays (cross 2)	χ^2 between Crosses	no. crosses	χ^2 between all Trays	no. trays
2	0.082	0.042	0.004	2	0.128	4
3	0.341	0.000	18.462* **	2	19.038* ** †	4
4	0.268	-	1 cross	1	0.268	2
5	0.967	1.176	8.357* **	2	10.304*	4
6	0.884	0.000	14.579* **	2	16.132* ** †	4
7	1 tray	1.923	28.719* **	2	29.127* ** †	4
8	0.073	0.260	2.256	2	2.555	4
9	4.32*	0.000	2.911	2	8.733*	4
10	0.508	2.237	7.188* **	2	8.681*	4
11	10.076* **	0.094	11.127* **	2	23.85* ** †	4
12	0.022	0.111	0.595	2	0.719	4
13	0.665	0.316	4.376*	2	5.492	4
14	0.000	0.067	0.033	2	0.099	4
15	0.257	1.094	2.599	2	3.504	4
16	1.052	4.286*	1.827	2	8.887*	4
17	4.286*	-	1 cross	1	4.286*	2
19	0.000	-	1 cross	1	0.000	2
21	0.800	0.001	0.069	2	1.302	4
22	0.000	0.000	0	2	0.000	4
24	4.464*	0.259	7.661* **	2	11.658* **	4
27	0.000	2.521	7.342* **	2	12.416* **	4
29	1 tray	0.317	3.008	2	3.369	3
33	0.098	0.483	0.368	2	0.914	4
35	0.077	-	1 cross	1	0.077	2
37	0.341	-	1 cross	1	0.341	2
41	0.000	-	1 cross	1	0.000	2
42	0.373	0.111	0.455	2	0.960	4
44	0.000	-	1 cross	1	0.000	2
45	1 tray	10.254* **	0.332	2	3.070	3
46	0.218	0.388	0.501	2	1.065	4
50	1.456	1.094	4.404*	2	7.039	4
52	0.183	0.162	0.334	2	0.675	4
55	0.111	16.652* **	9.195* **	2	28.43* ** †	4
56	0.259	1.032	0.486	2	1.843	4

Appendix 6.1: Chi-square values between replicates in the F_2 x Cop crosses

F_2	χ^2 between Trays (cross 1)	χ^2 between Trays (cross 2)	χ^2 between Crosses	no. crosses	χ^2 between all Trays	no. trays
57	6.966* **	0.487	2.473	2	10.503*	4
59	0.669	0.508	2.36	2	3.469	4
63	0.001	6.709* **	2.161	2	11.01*	4
64	3.872*	1 tray	0	2	3.893	3
65	0.884	0.078	1.836	2	2.682	4
66	0.741	1 tray	28.364* **	2	28.7* ** †	4
67	0.000	0.596	3.549	2	4.610	4
69	0.111	1.017	9.259* **	2	9.63*	4
71	1.456	0.063	0.03	2	1.498	4
72	5.455*	0.000	0.12	2	6.126	4
73	0.131	-	1 cross	1	0.131	2
75	2.069	0.059	9.918* **	2	10.608*	4
76	0.000	1.017	0.941	2	2.892	4
78	0.000	0.000	2.034	2	2.034	4
79	1 tray	0.111	1.182	2	1.280	3
80	0.000	-	1 cross	1	0.000	2
81	0.077	2.014	0.51	2	2.675	4
82	0.113	0.000	4.332*	2	4.491	4
83	0.000	0.000	2.034	2	2.034	4
85	0.000	0.218	3.651	2	3.787	4
88	0.287	2.411	1.234	2	3.977	4
89	0.041	3.265	0.281	2	3.688	4
90	0.381	0.000	2.565	2	3.030	4
91	1 tray	0.217	0.539	2	0.771	3
92	0.136	0.260	0.581	2	0.980	4
93	0.000	0.000	0	2	0.000	4
94	1.200	0.180	4.382*	2	5.927	4
95	0.000	0.000	2.035	2	2.035	4
96	0.113	17.33* **	2.184	2	21.835* ** †	4
97	4.296*	0.183	0.103	2	4.188	4
98	0.486	0.230	0.934	2	4.153	4
99	13.605* **	1.782	8.344* **	2	25.254* ** †	4
100	0.218	0.000 1.017	3.75	3	4.286	6
101	0.052	0.037	1.362	2	1.451	4
102	0.170	0.050	5.9*	2	6.125	4
103	3.835	0.136 3.136	1.769	3	8.577	6
104	1 tray	0.073	8.439* **	2	8.527*	3
105	0.012	0.821	0.025	2	0.863	4
106	0.031	-	1 cross	1	0.031	2
107	0.160	-	1 cross	1	0.160	2
108	0.843	0.100	5.394*	2	4.887	4
109	2.449	0.141	0.466	2	2.981	4
110	0.483	0.111	0.024	2	0.609	4
111	0.259	0.000	5.05*	2	5.537	4
112	0.070	2.297	0.138	2	3.160	4
113	0.162	-	1 cross	1	0.162	2
114	0.050	0.000	7.801* **	2	7.871*	4
115	0.000	4.164*	11.83* **	2	19.543* ** †	4
116	0.148	0.287	4.46*	2	4.875	4
117	0.000	0.315	0.31	2	0.701	4
118	0.000	0.287	0.142	2	0.450	4
119	0.278	1.364	0.006	2	1.651	4
120	0.162	0.827	19.811* **	2	20.869* ** †	4

Appendix 6.1: Chi-square values between replicates in the F_2 x Cop crosses

F_2	χ^2 between Trays (cross 1)	χ^2 between Trays (cross 2)	χ^2 between Crosses	no. crosses	χ^2 between all Trays	no. trays
121	0.000	0.000	0	2	0.000	4
122	0.218	3.791	49.368* **	2	52.592* ** †	4
123	0.315	0.404	4.274*	2	4.801	4
124	0.214	0.747	0	2	1.493	4
125	0.558	0.113	0	2	0.671	4
126	1.469	8.661* **	0.058	2	10.461*	4
127	0.982	0.006	3.059	2	4.144	4
129	0.002	0.483	0.931	2	1.501	4
131	1.074	2.541	0.086	2	3.538	4
132	4.975*	2.559	11.392* **	2	17.45* ** †	4
133	0.202	1.634	4.553*	2	6.617	4
134	0.422	0.111	5.927*	2	6.503	4
135	0.397	0.739	9.407* **	2	10.571*	4
136	0.073	0.912	1.704	2	2.796	4
137	0.001	0.659	19.064* **	2	20.026* ** †	4
138	3.519	-	1 cross	1	3.519	2
139	0.000	0.050	0.026	2	0.077	4
140	0.006	0.000	4.294*	2	4.304	4
141	1.351	7.937* **	5.032*	2	15.711* ** †	4
142	0.000	2.002	2.317	2	6.504	4
143	1.017	2.500	26.728* **	2	30.467* ** †	4
144	2.845	1 tray	15.94* **	2	16.95* ** †	3
145	0.025	0.884	0.409	2	1.592	4
146	0.000	3.060	3.084	2	9.032*	4
147	0.131	0.043	3.566	2	3.751	4
148	0.461	0.000	0.069	2	0.534	4
149	0.000	0.000	0	2	0.000	4
150	0.599	1.018	1.245	2	2.662	4
151	0.082	0.922	2.436	2	3.496	4
152	0.157	-	1 cross	1	0.157	2
153	1.032	1.148	2.566	2	4.116	4
154	1.501	0.001	0.027	2	2.671	4
155	3.060	2.320 0.000	5.043	3	11.25*	6
156	0.884	0.000	14.579* **	2	16.132* ** †	4
157	0.310	0.041	7.928* **	2	8.143*	4
158	0.306	0.000	11.07* **	2	11.659* **	4
159	0.140	0.062	2.021	2	2.260	4
160	0.000	0.073	12.781* **	2	12.915* **	4
161	0.004	1.017	9.476* **	2	9.673*	4
162	0.924	0.000	1.763	2	2.616	4
163	0.030	0.004	5.361*	2	5.399	4
164	2.951	1.017	1.185	2	6.180	4
165	0.554	0.022	3.601	2	4.100	4
166	0.000	0.826	6.549*	2	8.135*	4
167	0.635	0.030	0.028	2	0.697	4
168	0.008	0.131 0.182	1.106	3	1.423	6
169	0.074	1 tray	0.558	2	0.639	3
170	0.000	0.001	2.034	2	2.036	4
171	1.038	0.094	1.653	2	2.991	4
172	0.003	1.854	2.691	2	4.866	4
173	0.094	0.147	1.839	2	2.088	4
174	0.004	1.023	0.051	2	1.108	4
175	0.000	0.002	0.455	2	0.457	4

F_2	χ^2 between Trays (cross 1)	χ^2 between Trays (cross 2)	χ^2 between Crosses	no. crosses	χ^2 between all Trays	no. trays
176	1 tray	0.094	6.429*	2	6.493*	3
177	2.624	0.020	7.103* **	2	10.482*	4
178	0.662	0.268	9.853* **	2	10.617*	4
179	3.869*	0.000	3.862*	2	6.806	4
180	3.354	2.908	0.678	2	6.898	4
181	14.117* **	0.012 0.049	5.944	3	16.74* **	6
182	3.280	2.018	3.329	2	8.043*	4
183	1.017	3.539	11.123* **	2	17.247* ** †	4
184	0.741	2.299	8.655* **	2	12.117* **	4
185	0.000	0.180	6.06*	2	6.428	4
187	0.617	0.041	1.781	2	2.471	4
188	2.052	1.326	0.327	2	3.736	4
189	0.144	0.151	6.231*	2	6.509	4
190	3.354	0.404	0.601	2	4.639	4
191	0.960	0.169	0.463	2	1.571	4
192	0.820	1 tray	0.07	2	0.908	3
193	1.191	0.000	2.374	2	3.965	4
194	0.316	4.248*	1.604	2	7.443	4
195	0.082	0.259	7.709* **	2	8.045*	4
196	1.681	1.809	0.434	2	3.922	4
197	0.000	2.174	11.499* **	2	14.031* ** †	4
198	1.002	1.331	8.941* **	2	11.184*	4
199	0.078	0.000	6.852* **	2	6.922	4
200	2.074	0.000	0.005	2	2.150	4
201	0.475	-	1 cross	1	0.475	2
202	0.429	0.000	14.902* **	2	15.688* ** †	4
203	2.069	1.017	0.342	2	3.761	4
204	0.000	0.000	2.034	2	0.000	4
Controls						
Cer 10 x Cop3a	0.271	-	0.000	2	2.715	4
Cer 10 x Cop3b	2.443	-	-	-	-	-

Appendix 6.2

Single Factor/ One Way ANOVA results for each of the three markers in Cerig that co-segregate with the hybrid necrosis phenotype

Marker e217

Genotypes:

Column 1: 202/202

Column 2: 202/205

Column 3: 205/205

Groups	*Count*	*Sum*	*Average*	*Variance*
Column 1	8	218.8405	27.35506	189.0681
Column 2	79	2419.538	30.62707	98.29697
Column 3	51	700.2734	13.73085	66.96704

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	8932.656	2	4466.328	48.86576	**1.08E-16**	**3.063204**
Within Groups	12338.99	135	91.39995			
Total	21271.65	137				

Marker e416

Genotypes:

Column 1: 233/233

Column 2: 233/2442

Column 3: 242/242

Groups	*Count*	*Sum*	*Average*	*Variance*
Column 1	11	310.1577	28.19615	148.5205
Column 2	76	2375.26	31.25342	81.25575
Column 3	50	645.9146	12.91829	62.99561

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	10318.71	2	5159.355	64.8174	**2.04E-20**	**3.063715**
Within Groups	10666.17	134	79.59829			
Total	20984.88	136				

Marker e279

Genotypes:

Column 1: 761/761

Column 2: 761/787

No 782/782 genotype

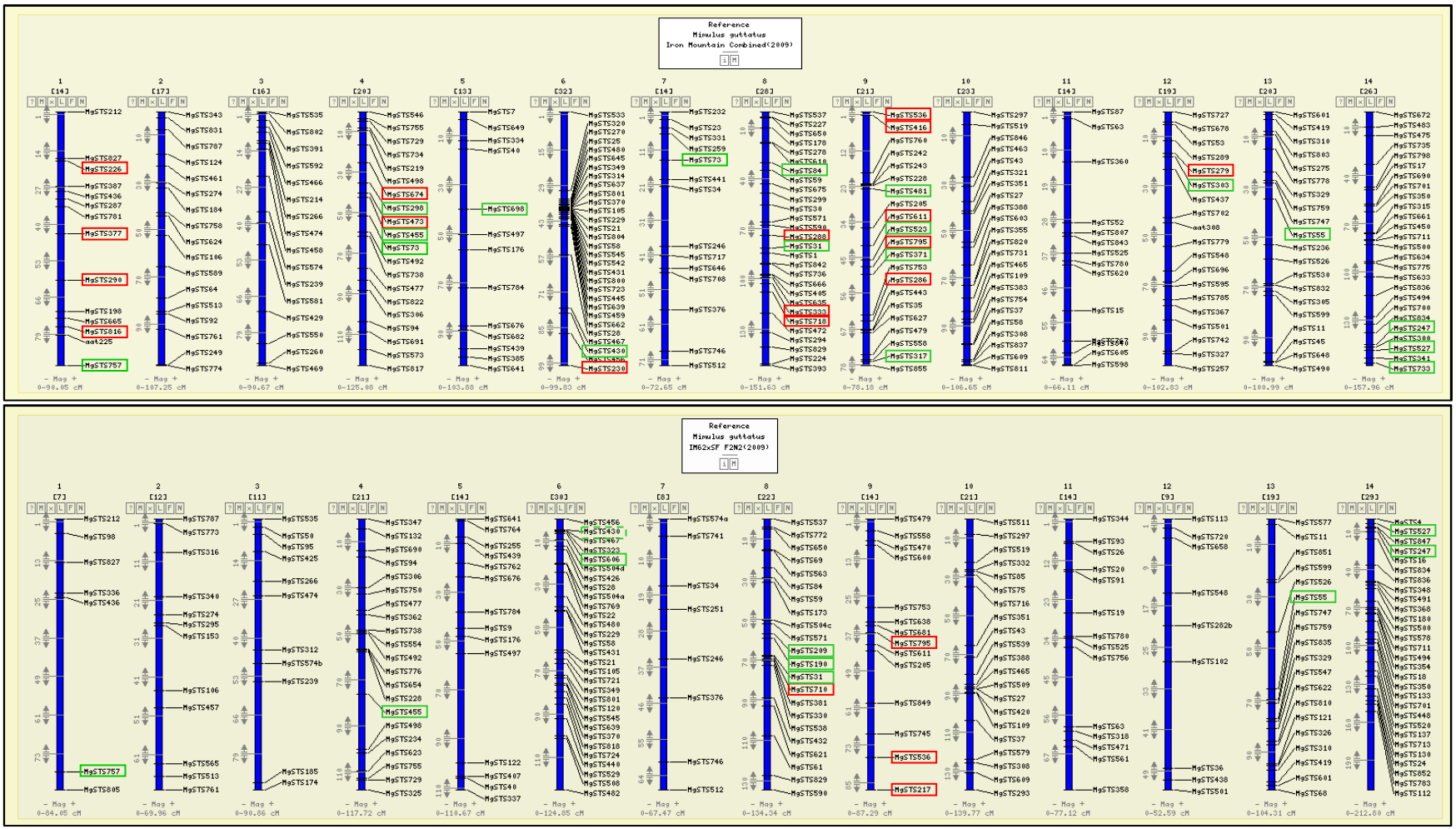
Groups	*Count*	*Sum*	*Average*	*Variance*
Column 1	30	884.5059	29.48353	185.1477
Column 2	79	1850.16	23.41975	141.9821

ANOVA

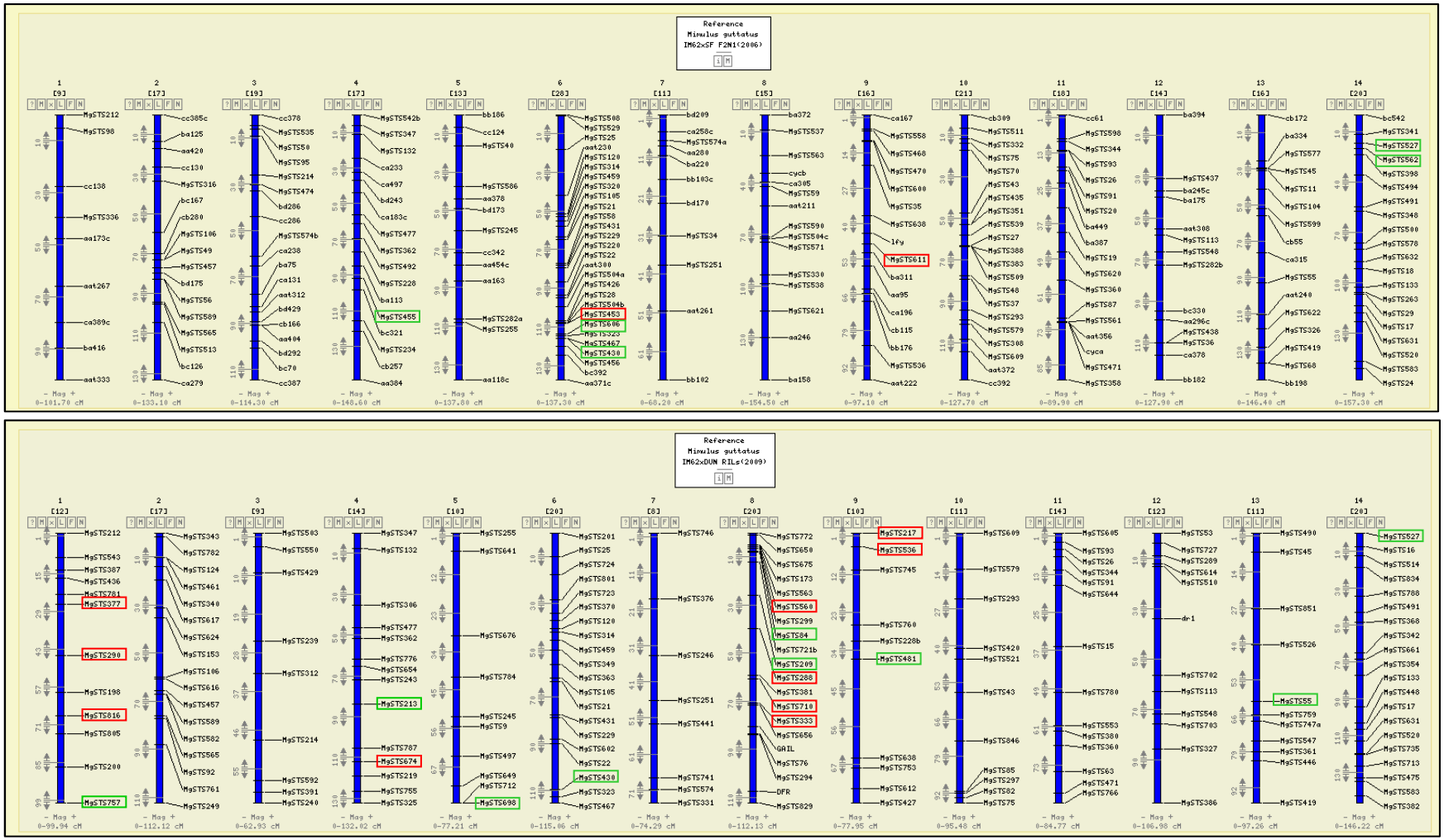
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	799.4824	1	799.4824	5.202214	**0.024539**	**3.929844**
Within Groups	16443.89	107	153.6812			
Total	17243.37	108				

Appendix 7.1

Figure A7.1 Maps (from www.mimulusevolution.com) of the *Mimulus* genome showing the size (cM) of the fourteen chromosomes and the locations of the MgSTS markers screened



Appendix 7.1: Maps of the *Mimulus* genome showing the size (cM) of the fourteen chromosomes and the locations of the MgSTS markers screened



All four maps are provided as no single map includes all the markers screened. The markers are outlined according to whether they showed variation (red) between Cer10 and Cer35 or had no variation (green) [NB. twenty-seven markers showed variation, but seven are not mapped]

References

ABBOTT, R. J. & COMES, H. P. (2007) Blowin' in the wind - the transition from ecotype to species. *New Phytologist,* 175: 197-200.

ALCAZAR, R., GARCIA, A. V., PARKER, J. E. & REYMOND, M. (2009) Incremental steps towards incompatibility revealed by *Arabidopsis* epistatic interactions modulating salicylic acid pathway activation. *Proceedings of the Natural Academy of Science,* 106: 334-339.

ALLEN, W. R. & SHEPPARD, P. M. (1971) Copper tolerance in some Californian populations of the monkey flower, *Mimulus guttatus*. *Proceedings of the Royal Society London,* 177: 177-196.

ALVAREZ, M. E. (2000) Salicylic acid in the machinery of hypersensitive cell death and disease resistance. *Plant Molecular Biology,* 44: 429-442.

ANTONOVICS, J., BRADSHAW, A. D. & TURNER, R. G. (1971) Heavy metal tolerance in plants. *Advances in Ecological Research* 7: 1-85.

BALCIUNAS, D. & RONNE, H. (2000) Evidence of domain swapping within the jumonji family of transcription factors. *Trends in Biochemical Sciences,* 25: 274-276.

BARBASH, D. A. & ASHBURNER, M. (2003) A novel system of fertility rescue in *Drosophila* hybrids reveals a link between hybrid lethality and female sterility. *Genetics,* 163: 217-226.

BARBASH, D. A., AWADALLA, P. & TARONE, A. M. (2004) Functional divergence caused by ancient positive selection of a *Drosophila* hybrid incompatibility locus. *PLoS Biology,* 2: 839-848.

BARBASH, D. A., ROOTE, J. & ASHBURNER, M. (2000) The *Drosophila melanogaster Hybrid male rescue* gene causes inviability in male and female species hybrids. *Genetics,* 154: 1747-1771.

BARBASH, D. A., SIINO, D. F., TARONE, A. M. & ROOTE, J. (2003) A rapidly evolving MYB-related protein causes species isolation in *Drosophila. PNAS,* 100: 5302-5307.

BARR, C. M. & FISHMAN, L. (2010) The nuclear component of a cytonuclear incompatibility in *Mimulus* maps to a cluster of pentatricopeptide repeat genes. *Genetics,* 184: 455-465.

BAYES, J. J. & MALIK, H. S. (2009) Altered heterochromatin binding by a hybrid sterility protein in *Drosophila* sibling species. *Science,* 326: 1538-1541.

BELKHADIR, Y., NIMCHUK, Z., HUBERT, D. A., MACKEY, D. & DANGL, J. L. (2004) *Arabidopsis* RIN4 negatively regulates disease resistance mediated by RPS2 and RPM1 downstream or independent of the NDR1 signal modulator and is not required for the virulence functions of bacterial type III effectors AvrRpt2 or AvrRpm1. *The Plant Cell,* 16: 2822-2835.

BENT, A. F. (1996) Plant disease resistance genes: function meets structure. *The Plant Cell,* 8: 1757-1771.

BIKARD, D., PATEL, D., LE METTE, C., GIORGI, V., CAMILLERI, C., BENNETT, M. J. & LOUDET, O. (2009) Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. *Science,* 323: 623-626.

BOLKAN, B. J., BOOKER, R., GOLDBERG, M. L. & BARBASH, D. A. (2007) Developmental and cell cycle progression defects in *Drosophila* hybrid males. *Genetics,* 177: 2233-2241.

BOMBLIES, K. & WEIGEL, D. (2007) Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. *Nature Reviews Genetics*, 8: 382-393.

BOMBLIES, K. (2006) Hybrid incompatibility: When opposites attract with a fatal outcome. *Current Biology,* 16: 1322-1328.

BOMBLIES, K. (2009) Too much of a good thing? Hybrid necrosis as a by-product of plant immune system diversification. *Botany,* 87: 1013-1022.

BOMBLIES, K., LEMPE, J., EPPLE, P., WARTHMANN, N., LANZ, C., DANGL, J. & WEIGEL, D. (2007) Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. *PLoS Biology,* 5: 1962-1972.

BRIDEAU, N. J., FLORES, H. A., WANG, J., MAHESHWARI, S., WANG, X. & BARBASH, D. A. (2006) Two Dobzhansky-Muller genes interact to cause hybrid lethality in *Drosophila*. *Science,* 314: 1292-1295.

BURT, A. & TRIVERS, R. (2006) *Genes in conflcit: the biology of selfish genetic elements*, Belknap Press, Harvard.

BURTON, T. L. & HUSBAND, B. C. (2000) Fitness differences among diploids, tetraploids, and their triploid progeny in *Chamerion angustifolium*: Mechanisms of inviability and implications for polyploid evolution. *Evolution,* 54: 1182-1191.

CASE, A. L. & WILLIS, J. H. (2008) Hybrid male sterility in *Mimulus* (Phrymaceae) is associated with a geographically restricted mitochondrial rearrangement. *Evolution,* 62: 1026-1039.

CHEN, J., DING, J., OUYANG, Y., DU, H., YANG, J., CHENG, K., ZHAO, J., QIU, S., ZHANG, X., YAO, J., LIU, K., WANG, L., XU, C., LI, X., XUE, Y., XIA, M., JI, Q., LU, J., XU, M. & ZHANG, Q. (2008) A triallelic system of *S5* is a major regulator of the reproductive barrier and compatibility of *indica-japonica* hybrids in rice. *Proceedings of the National Academy of Sciences,* 105: 11436-11441.

CHOU, J.-Y., HUNG, Y.-S., LIN, K.-H., LEE, H.-Y. & LEU, J.-Y. (2010) Multiple molecular mechanisms cause reproductive isolation between three yeast species. *PLoS Biology,* 8: e1000432.

CHRISTIE, P. & MACNAIR, M. R. (1984) Complementary lethal factors in two North American populations of the yellow monkey flower. *The Journal of Heredity,* 75: 510-511.

CHRISTIE, P. & MACNAIR, M. R. (1987) The distribution of postmating reproductive isolating genes in populations of the yellow monkey flower, *Mimulus guttatus*. *Evolution,* 41: 571-578.

CHU, C.-G., FARIS, J. D., FRIESEN, T. L. & XU, S. S. (2006) Molecular mapping of hybrid necrosis genes *Ne1* and *Ne2* in hexaploid wheat using microsatellite markers. *Theoretical Applied Genetics,* 112: 1374-1381.

COYNE, J. A. & ORR, H. A. (1989) Patterns of speciation in *Drosophila*. *Evolution,* 43: 362-381.

COYNE, J. A. & ORR, H. A. (1998) The evolutionary genetics of speciation. *Philosophical Transactions of the Royal Society London,* 353: 287-305.

COYNE, J. A. & ORR, H. A. (2004) *Speciation.*, Sinauer Associates, Inc., Sunderland, Massachusetts, U.S.A.

DANGL, J. L. & JONES, J. D. G. (2001) Plant pathogens and integrated defence responses to infection. *Nature,* 411: 826-833.

DANGL, J. L., DIETRICH, R. A. & RICHBERG, M. H. (1996) Death don't have no mercy: cell death programs in plant-microbe interactions. *The Plant Cell,* 8: 1793-1807.

DARWIN, C. (1859) *The origin of species by means of natural selection.* John Murray, London. Reprinted 1985 by Penguin Books Ltd, London. Edited with an introduction by J.W. Burrow.

DE QUEIROZ, K. (1998) The general lineage concept of species, species criteria, and the process of speciation. In Howard, D.J. & Berlocher, S.H. (Eds.) *Endless Forms: Species and Speciation.* Oxford University Press.

DETTMAN, J. R., ANDERSON, J. B. & KOHN, L. M. (2008) Divergent adaptation promotes reproductive isolation among experimental populations of the filamentous fungus *Neurospora*. *Evolutionary Biology,* 8:35 (doi:10.1186/1471-2148-8-35)

DETTMAN, J. R., ANDERSON, J. B. & KOHN, L. M. (2009) Genome-wide investigation of reproductive isolation in experimental lineages and natural species of *Neurospora*: identifying candidate regions by microarray-based genotyping and mapping. *Evolution,* 64: 694-709.

DING, J., ZHANG, W. L., JING, Z. Q., CHEN, J. Q. & TIAN, D. C. (2007) Unique pattern of *R* gene variation within populations in *Arabidopsis*. *Molecular Genetics and Genomics,* 277: 619-629.

DOBZHANSKY, T. (1936) Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. *Genetics,* 21: 113-135.

DOBZHANSKY, T. (1937) *Genetics and the origin of species.* Columbia University Press, New York.

DOYLE, J. J. & DOYLE, J. L. (1990) Isolation of plant DNA from fresh tissue. *Focus,* 12: 13-15.

DUNN, L. C. (1956) Evidence of evolutionary forces leading to the spread of lethal genes in wild populations of house mice. *Genetics,* 43: 158-163.

EDREVA, A. (2005) Pathogenesis-related proteins: research progress in the last 15 years. *General and Applied Plant Physiology,* 31: 105-124.

FAY, J.C. & WU, C.I. (2000) Hitchhiking under positive Darwinian selection. *Genetics*, 155: 1405-1413.

FERREE, P. M. & BARBASH, D. A. (2009) Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid sterility in *Drosophila*. *PLoS Biology,* 7: e1000234.

FISHMAN, L. & WILLIS, J. H. (2001) Evidence for Dobzhansky-Muller incompatibilities contributing to the sterility of hybrids between *Mimulus guttatus* and *M. nasutus*. *Evolution,* 55: 1932-1942.

FISHMAN, L. & WILLIS, J. H. (2006) A cytonuclear incompatibility causes anther sterility in *Mimulus* hybrids. *Evolution,* 60: 1372-1381.

FITZPATRICK, B. M. (2008) Hybrid dysfunction: population genetic and quantitative genetic perspectives. *The American Naturalist,* 171: 491-498.

FLINT, J. & MOTT, R. (2001) Finding the molecular basis of quantitative traits: successes and pitfalls. *Nature Reviews Genetics,* 2: 437-445.

FUKUOKA, S., NAMAI, H. & OKUNO, K. (1998) RFLP mapping of the genes controlling hybrid breakdown in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics,* 97: 446-449.

FUKUOKA, S., NEWINGHAM, M. C. V., ISHTIAQ, M., NAGAMINE, T., KAWASE, M. & OKUNO, K. (2005) Identification and mapping of two new loci for hybrid breakdown in cultivated rice. *Rice Genetics Newsletter,* 22: 29-30.

GACHON, C. M. M., LANGLOIS-MEURINNE, M. & SAINDRENAN, P. (2005) Plant secondary metabolism glycosyltransferases: the emerging functional analysis. *Trends in Plant Science,* 10: 542-549.

GREENBERG, J. T. (1997) Programmed cell death in plant-pathogen interactions. *Annual Review of Plant Physiology and Plant Molecular Biology,* 48: 525-545.

GREENSHIELDS, D. L. & JONES, J. D. G. (2008) Plant pathogen effectors: getting mixed messages. *Current Biology,* 18: R128-R130.

HAILING, J., LIU, Y., YANG, K.-Y., KIM, C. Y., BAKER, B. & ZHANG, S. (2003) Function of a mitogen-activated protein kinase pathway in *N* gene-mediated resistance in tobacco. *The Plant Journal,* 33: 719-731.

HANNAH, M. A., KRAMER, K. M., GEOFFROY, V., KOPKA, J., BLAIR, M. W., ERBAN, A., VALLEJOS, C. E., HEYER, A. G., SANDERS, F. E. T., MILLNER, P. A. & PILBEAM, D. J. (2007) Hybrid weakness controlled by the dosage-dependent lethal (DL) gene system in common bean (*Phaseolus vulgaris*) is caused by a shoot-derived inhibitory signal leading to salicylic acid-associated root death. *New Phytologist,* 176: 537-549.

HARPER, F. A., SMITH, S. E. & MACNAIR, M. R. (1997) Where is the cost in copper tolerance in *Mimulus guttatus*? Testing the trade-off hypothesis. *Functional Ecology,* 11: 764-774.

HARRISON, J. S. & BURTON, R. S. (2006) Tracing hybrid incompatibilities to single amino acid substitutions. *Molecular Biology and Evolution,* 23: 559-564.

HAWEKER, H., RIPS, S., KOIWA, H., SALOMON, S., SAIJO, Y., CHINCHILLA, D., ROBATZEK, S. & VON SCHAEWEN, A. (2010) Pattern recognition receptors require *N*-glycosylation to mediate plant immunity. *The Journal of Biological Chemistry,* 285: 4629-4636.

HERBERS, K., MEUWLY, P., FROMMER, W. B., METRAUX, J.-P. & SONNEWALD, U. (1996) Systemic acquired resistance mediated by the ectopic expression of invertase: possible hexose sensing in the secretory pathway. *The Plant Cell,* 8: 793-803.

HERMSEN, J. G. T. (1963) The genetic basis of hybrid necrosis in wheat. *Genetica,* 33: 245-287.

HOLLINGSHEAD, L. (1930) A lethal factor in *Crepis* effective only in an interspecific hybrid. *Genetics,* 15: 114-140.

HUTTER, P. & ASHBURNER, M. (1987) Genetic rescue of inviable hybrids between *Drosophila melanogaster* and its sibling species. *Nature,* 327: 331-333.

HUTTER, P., ROOTE, J. & ASHBURNER, M. (1990) A genetic basis for the inviability of hybrids between sibling species of *Drosophila*. *Genetics,* 124: 909-920.

ICHITANI, K., FUKUTA, Y., TAURA, S. & SATO, M. (2001) Chromosomal location of *Hwc2*, one of the complementary hybrid weakness genes, in rice. *Plant Breeding,* 120: 523-525.

ICHITANI, K., NAMIGOSHI, K., SATO, M., TAURA, S., AOKI, M., MATSUMOTO, Y., SAITOU, T., MARUBASHI, W. & KUBOYAMA, T. (2007) Fine mapping and allelic dosage effect of *Hwc1*, a complementary hybrid weakness gene in rice. *Theoretical and Applied Genetics,* 114: 1407-1415.

INOUE, E., MARUBASHI, W. & NIWA, M. (1996) Genomic factors controlling the lethality exhibited in the hybrid between *Nicotiana suaveolens* Lehm. and *N. tabacum* L. *Theoretical and Applied Genetics,* 93: 341-347.

ISPOLATOV, I. & DOEBELI, M. (2009) Speciation due to hybrid necrosis in plant-pathogen models. *Evolution,* 63: 3076-3084.

JEUKEN, M. & LINDHOUT, P. (2002) *Lactuca saligna*, a non-host for lettuce downy mildew (*Bremia lactucae*), harbors a new race-specific *Dm* gene and three QTLs for resistance. *Theoretical and Applied Genetics,* 105: 384-391.

JEUKEN, M. J. W., ZHANG, N. W., MCHALE, L. K., PELGROM, K., DEN BOER, E., LINDHOUT, P., MICHELMORE, R. W., VISSER, R. G. F. & NIKS, R. E. (2009) *Rin4* causes hybrid necrosis and race-specific resistance in an interspecific lettuce hybrid. *The Plant Cell,* 21: 3368-3378.

JIANG Q, CHEN, H., PAN, X., PAN, Q., SHI, Y., LI, X., ZHANG, G., WANG, Y., XIE, S. & SHEN, S. (2008) Proteomic analysis of wheat (*Triticum aestivum* L.) hybrid necrosis. *Plant Science,* 175: 394-401.

JIANG W, CHU, S.-H., PIAO, R., CHIN, J.-H., JIN, Y.-M., LEE, J., QIAO, Y., HAN, L., PIAO, Z. & KOH, H.-J. (2008) Fine mapping and candidate gene analysis of *hwh1* and *hwh2*, a set of complementary genes controlling hybrid breakdown in rice. *Theoretical and Applied Genetics,* 116: 1117-1127.

JOHNSON, N. A. (2010) Hybrid incompatibility genes: remnants of a genomic battlefield? *Trends in Genetics,* 26: 317-325.

JONES, J. D. G. & DANGL, J. L. (2006) The plant immune system. *Nature,* 444: 323-329.

KEEGSTRA, K. & RAIKHEL, N. (2001) Plant glycosyltransferases. *Current Opinion in Plant Biology,* 4: 219-224.

KHANNA-CHOPRA, R., DALAL, M., KUMAR, G. P. & LALORAYA, M. (1998) A genetic system involving superoxide causes F_1 necrosis in wheat (*T. aestivum* L.). *Biochemical and Biophysical Research Communications,* 248: 712-715.

KLESSIG, D. F. & MALAMY, J. (1994) The salicylic acid signal in plants. *Plant Molecular Biology,* 26: 1439-1458.

KLOSE, R. J., KALLIN, E. M. & ZHANG, Y. (2006) JmjC-domain-containing proteins and histone demethylation. *Nature Reviews Genetics,* 7: 715-727.

KOSTOFF, D. (1930) Ontogeny, genetics and cytology of *Nicotiana* hybrids. *Genetica,* 12: 33-139.

KRUGER, J., THOMAS, C. M., GOLSTEIN, C., DIXON, M. S., SMOKER, M., TANG, S., MULDER, L. & JONES, J. D. G. (2002) A tomato cysteine protease required for *Cf-2*-dependent disease resistance and supression of autonecrosis. *Science,* 296: 744-747.

KUBO, T. & YOSHIMURA, A. (2002) Genetic basis of hybrid breakdown in a *Japonica/Indica* cross of rice, *O. sativa* L. *Theoretical and Applied Genetics,* 105: 906-911.

KUBO, T. & YOSHIMURA, A. (2005) Epistasis underlying female sterility detected in hybrid breakdown in a Japonica-Indica cross of rice (*Oryza sativa* L.). *Theoretical and Applied Genetics,* 110: 346-355.

LAMB, C. & DIXON, R. A. (1997) The oxidative burst in plant disease resistance. *Annual Review Plant Physiology and Plant Molecular Biology,* 48: 251-275.

LEE, H.-Y., CHOU, J.-Y., CHEONG, L., CHANG, N.-H., YANG, S.-Y. & LEU, J.-Y. (2008) Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. *Cell,* 135: 1065-1073.

LEE, J. A. (1981) Genetics of D_3 complementary lethality in *Gossypium hirsutum* and *G. barbadense*. *The Journal of Heredity,* 72: 299-300.

LIM, E.-K. & BOWLES, D. J. (2004) A class of glycosyltransferases involved in cellular homeostasis. *The EMBO Journal,* 23: 2915-2922.

LIM, E.-K., DOUCET, C. J., LI, Y., ELIAS, L., WORRALL, D., SPENCER, S. P., ROSS, J. & BOWLES, D. J. (2002) The activity of *Arabidopsis* glycosyltransferases toward salicylic acid, 4-hydrobenzoic acid, and other benzoates. *The Journal of Biological Chemistry,* 277: 586-592.

LONG, Y., ZHAO, L., NIU, B., SU, J., WU, H., CHEN, Y., ZHANG, Q., GUO, J., ZHUANG, C., MEI, M., XIA, J., WANG, L., WU, H. & LIU, Y.-G. (2008) Hybrid male sterility in rice controlled by interaction between divergent alleles of two adjacent genes. *Proceedings of the National Academy of Sciences,* 105: 18871-18876.

LOWRY, D. B., MODLISZEWSKI, J. L., WRIGHT, K. M., WU, C. A. & WILLIS, J. H. (2008) The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philosophical Transactions Of The Royal Society,* 363: 3009-3021.

MACNAIR, M. R. & CHRISTIE, P. (1983) Reproductive isolation as a pleiotropic effect of copper tolerance in *Mimulus guttatus*? *Heredity,* 50: 295-302.

MACNAIR, M. R. & GARDNER, M. (1998) The evolution of edaphic endemics. In Howard, D.J. & Berlocher, S.H. (Eds.) *Endless Forms: Species and Speciation.* Oxford University Press.

MACNAIR, M. R. & WATKINS, A. D. (1983) The fitness of the copper tolerance gene in *Mimulus guttatus* in uncontaminated soil. *New Phytologist,* 95: 133-137.

MACNAIR, M. R. (1981) The uptake of copper by plants of *Mimulus guttatus* differing in genotype primarily at a single major copper tolerance locus. *New Phytologist,* 88: 723-730.

MACNAIR, M. R. (1983) The genetic control of copper tolerance in the yellow monkey flower, *Mimulus guttatus*. *Heredity,* 50: 283-293.

MACNAIR, M. R. (1987a) A model for the spread of genes for post-mating reproductive isolation by natural selection. *Journal of Theoretical Biology,* 125: 105-115.

MACNAIR, M. R. (1987b) Heavy metal tolerance in plants: A model evolutionary system. *Trends in Ecology & Evolution,* 2: 354-359.

MACNAIR, M. R. (1992) Preliminary studies on the genetics and evolution of the serpentine endemic *Mimulus nudatus* Curran. In Baker, A.J.M., Proctor, J. & Reeves, R. D. (Eds.) *The vegetation of ultramafic (serpentine) soils.* Intercept Ltd, England.

MACNAIR, M. R. (1993) A polymorphism for a lethal phenotype governed by two duplicate genes in *Mimulus guttatus*. *Heredity,* 70: 362-369.

MACNAIR, M. R. (1997) The evolution of plants in metal-contaminated environments. In Bijlsma, R. & Loeschcke, V.(Eds.) *Environmental stress, adaptation and evolution.* Birkhauser.

MACNAIR, M. R., MACNAIR, V. E. & MARTIN, B. E. (1989) Adaptive speciation in *Mimulus*: an ecological comparison of *M. cupriphilus* with its presumed progenitor, *M. guttatus*. *New Phytologist,* 112: 269-279.

MACNAIR, M. R., SMITH, S. E. & CUMBES, Q. J. (1993) Heritability and distribution of variation in degree of copper tolerance in *Mimulus guttatus* at Copperopolis, California. *Heredity,* 71: 445-455.

MAHESHWARI, S., WANG, J. & BARBASH, D. A. (2008) Recurrent positive selection of the *Drosophila* hybrid incompatibility gene *Hmr*. *Molecular Biology and Evolution,* 25: 2421-2430.

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

MARUBASHI, W., YAMADA, T. & NIWA, M. (1999) Apoptosis detected in hybrids between *Nicotinana glutinosa* and *N. repanda* expressing lethality. *Planta,* 210: 168-171.

MASUDA, Y., YAMADA, T., KUBOYAMA, T. AND MARUBASHI, W. (2007) Identification and charaterization of genes involved in hybrid lethality in hybrid tobacco cells (*Nicotiana suaveolens* x *N. tabacum*) using supression subtractive hybridization. *Plant Cell Reports,* 26: 1595-1604.

MATSUBARA, K., ANDO, T., MIZUBAYASHI, T., ITO, S. & YANO, M. (2007) Identification and linkage mapping of complementary recessive genes causing hybrid breakdown in an intraspecific rice cross. *Theoretical and Applied Genetics,* 115: 179-186.

MATSUBARA, K., ITO, S., NONOUE, Y., ANDO, T. & YANO, M. (2006) A novel gene responsible for hybrid breakdown found in a cross between japonica and indica cultivars in rice. *Rice Genetics Newsletter,* 23: 11.

MAYR, E. (1942) *Systematics and the Origin of Species*. Columbia University Press, New York.

MAYR, E. (1963) Species concepts and their application. In Ridley, M. (Ed.) *Evolution*. Oxford University Press, 1997.

MAYR, E. (2001) Wu's genic view of speciation - Commentary. *Journal of Evolutionary Biology,* 14: 866-867.

MCDERMOTT, S. R. & NORR, M. A. F. (2010) The role of meiotic drive in hybrid male sterility. *Philosopohcal Transactions of the Royal Society B,* 365: 1265-1272.

MCDOWELL, J. M. & SIMON, S. A. (2006) Recent insights into *R* gene evolution. *Molecular Plant Pathology,* 7: 437-448.

MCHALE, L., TAN, X., KOEHL, P. & MICHELMORE, R. W. (2006) Plant NBS-LRR proteins: adaptable guards. *Genome Biology,* 7: 212 (doi:10.1186/gb-2006-7-4-212)

MESHI, T. & IWABUCHI, M. (1995) Plant transcription factors. *Plant Cell Physiology,* 36: 1405-1420.

MESTRES, F., BALANYA, J., PASCUAL, M., ARENAS, C., GILCHRIST, G. W., HUEY, R. B. & SERRA, L. (2009) Evolution of Chilean colonizing populations of *Drosophila subobscura*: lethal genes and chromosomal arrangements. *Genetica,* 136: 37-48.

MICHELMORE, R. W. & MEYERS, B. C. (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Research,* 8: 1113-1130.

MICHELMORE, R. W., PARAN, I. & KESSELI, R. V. (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations, *Genetics,* 88: 9828-9832.

MIHOLA, O., TRACHTULEC, Z., VLCEK, C., SCHIMENTI, J. C. & FOREJT, J. (2009) A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science,* 323: 373-375.

MINO, M., MAEKAWA, K., OGAWA, K., YAMAGISHI, H. & INOUE, M. (2002) Cell death processes during expression of hybrid lethality in interspecific F_1 hybrid between *Nicotiana gossei* domin and *Nicotiana tabacum*. *Plant Physiology,* 130: 1-12.

MIURA, K., YAMAMOTO, E., MORINAKA, Y., TAKASHI, T., KITANO, H., MATSUOKA, M. & ASHIKARI, M. (2008) The hybrid breakdown *1(t)* locus induces interspecific hybrid breakdown between rice *Oryza sativa* cv. Koshihikari and its wild relative *O. nivara*. *Breeding Science,* 58: 99-105.

MOYLE, L. C. & GRAHAM, E. B. (2005) Genetics of hybrid incompatibility between *Lycopersicon esculentum* and *L. hirsutum*. *Genetics,* 169: 355-373.

MULLER, H. J. (1942) Isolating mechanisms, evolution and temperature. *Biological Symposia,* 6: 71-125.

NOLTE, V., WIEGEL, D. & SCHLOTTERER, C. (2008) The impact of shared ancestral variation on hybrid male lethality - a 16 codon indel in the *Drosophila simulans Lhr* gene. *Journal of Evolutionary Biology,* 21: 551-555.

OKA, H.-I. (1956) Phylogenetic differentiation of cultivated rice. XV. Complementary lethal genes in rice. *The Japanese Journal of Genetics,* 32: 83-87.

OLIVER, P. L., GOODSTADT, L., BAYES, J. J., BIRTLE, Z., ROACH, K. C., PHADNIS, N., BEATSON, S. A., LUNTER, G., MALIK, H. S. & PONTING, C. P. (2009) Accelerated evolution of the *Prdm9* speciation gene across diverse metazoan taxa. *PLoS Genetics,* 5 (12): e1000753.

ORR, H. A. & IRVING, S. (2000) Genetic Analysis of the *Hybrid male rescue* locus of *Drosophila*. *Genetics,* 155: 225-231.

ORR, H. A. (1995) The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics,* 139: 1805-1813.

ORR, H. A., MADDEN, L. D., COYNE, J. A., GOODWIN, R. & HAWLEY, R. S. (1997) The developmental genetics of hybrid inviability: a mitotic defect in *Drosophila* hybrids. *Genetics,* 145: 1031-1040.

ORR, H. A., MASLY, J. P. & PRESGRAVES, D. C. (2004) Speciation genes. *Current Opinion in Genetics & Development,* 14: 675-679.

ORTIZ-BARRIENTOS, D. & KANE, N. C. (2007) Meeting review: American Genetics Association Symposium on the genetics of speciation. *Molecular Ecology,* 16: 2852-2854.

PATTERSON, H. E. H. (1985) The recognition species concept of species. In Vrba E.S. (Ed.) *Species and Speciation.* Transvaal Museum Monograph No. 4, Pretoria: 21-29.

PFLUGER, J. & WAGNER, D. (2007) Histone modifications and dynamic regulation of genome accessibility in plants. *Current Opinion in Plant Biology,* 10: 645-652.

PHADNIS, N. & ORR, H. A. (2009) A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids. *Science,* 323: 376-379.

PHILLIPS, L. L. (1977) Interspecific incompatibility in *Gossypium*. IV. Temperature-conditional lethality in hybrids of *G. klotzschianum*. *American Journal of Botany,* 64: 914-915.

PRESGRAVES, D. C. (2007) Speciation Genetics: Epistasis, conflict and the origin of species. *Current Biology,* 17: R125-R127.

PRESGRAVES, D. C. (2010) The molecular evolutionary basis of species formation. *Nature Reviews Genetics,* 11: 175-180.

PRESGRAVES, D. C., BALAGOPALAN, L., ABMAYR, S. M. & ORR, H. A. (2003) Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. *Nature,* 423: 715-719.

PRIGENT, S. R., MATSUBAYASHI, H. & YAMAMOTO, M.-T. (2009) Transgenic *Drosophila simulans* strains prove the identity of the speciation gene *Lethal hybrid rescue*. *Genes and Genetic Systems,* 84: 353-360.

QUARRIE, S. A., LAZIC-JANCIC, V., KOVACEVIC, D., STEED, A. & PEKIC, S. (1999) Bulk segregant analysis with molecular markers and its use for improving drought resistance in maize. *Journal of Experimental Botany,* 50: 1299-1306.

RAMSEY, J., BRADSHAW, H. D. J. & SCHEMSKE, D. W. (2003) Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phymaceae). *Evolution,* 57: 1520-1534.

RASKIN, I. (1992) Role of salicylic acid in plants. *Annual Review Plant Physiology and Plant Molecular Biology,* 43: 439-463.

RIESEBERG, L. H. & CARNEY, S. E. (1998) Tansley Review No. 102: Plant hybridization. *New Phytologist,* 140: 599-624.

RIESEBERG, L. H. & WILLIS, J. H. (2007) Plant Speciation. *Science,* 317: 910-914.

RIESEBERG, L. H. (2001) Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution,* 16: 351-358.

SAITO, T., ICHITANI, K., SUZUKI, T., MARUBASHI, W. & KUBOYAMA, T. (2007) Developmental observation and high temperature rescue from hybrid weakness in a cross between Japanese rice cultivars and Peruvian rice cultivar 'Jamaica'. *Breeding Science,* 57: 281-288.

SAMORA, P. J., STELLY, D. M. & KOHEL, R. J. (1994) Localization and mapping of the Le_1 and Gl_2 loci of cotton (*Gossypium hirsutum* L.). *The Journal of Heredity,* 85: 152-157.

SATO, Y.-I. & MORISHIMA, H. (1987) Studies on the distribution of complementary gene causing F_1 weakness in common rice and its wild relatives. 2. Distribution of two complementary genes, *Hwc1* and *Hwc2*, gene in native cultivars and its wild relative of tropical Asia. *Euphytica,* 36: 425-431.

SAWAMURA, K. (2000) Genetics of hybrid inviability and sterility in *Drosophila*: The *Drosophila melanogaster-Drosophila simulans* case. *Plant Species Biology,* 15: 237-247.

SAWAMURA, K., TAIRA, T. & WATANABE, T. K. (1993a) Hybrid lethal systems in the *Drosophila melanogaster* species complex I. The *maternal hybrid rescue* (*mhr*) gene of *Drosophila simulans*. *Genetica,* 133: 299-305.

SAWAMURA, K., WATANABE, T. K. & YAMAMOTO, M.-T. (1993b) Hybrid lethal systems in the *Drosophila melanogaster* species complex. *Genetica,* 88: 175-185.

SAWAMURA, K., YAMAMOTO, M.-T. & WATANABE, T. K. (1993c) Hybrid lethal systems in the *Drosophila melanogaster* species complex II.The *zygotic hybrid rescue* (*zhr*) gene of *D. melanogaster*. *Genetica,* 133: 307-313.

SCHARTL, M. (2008) Evolution of *Xmrk*: an oncogene, but also a speciation gene? *Bioessays,* 30: 822-832.

SCHUELKE, M. (2000) An economic method for the fluorescent labelling of PCR fragments. *Nature Biotechnology*, 18: 233-234

SCOPECE, G., WIDMER, A. & COZZOLINO, S. (2008) Evolution of postzygotic reproductive isolation in a guild of deceptive orchids. *The Amercian Naturalist,* 171: 315-326.

SINGH, S., CHAUDHARY, H. K. & SETHI, G. S. (2000) Distribution and allelic expressivity of genes for hybrid necrosis in some elite winter and spring wheat ecotypes. *Euphytica,* 112: 95-100.

SOBRIZAL & YOSHIMURA, A. (2009) Hybrid weakness found in the backcross progeny of an interspecific cross of *Oryza sativa* /*O. glumaepatula*. *Atom Indonesia,* 35: 49-56.

SONG, L., GUO, W. & ZHANG, T. (2009) Interaction of novel Dobzhansky-Muller type genes for the induction of hybrid lethality between *Gossypium hirsutum* and *G. barbadense* cv. Coastland R4-4. *Theoretical Applied Genetics,* 119: 33-41.

STELLY, D. M. (1990) Localization of the Le_2 locus of cotton (*Gossypium hirsutum* L.). *Journal of Heredity,* 81: 193-197.

STURTEVANT, A. H. (1920) Genetic studies on *Drosophila simulans*. I. Introduction. Hybrids with *Drosophila melanogaster*. *Genetics,* 5: 488-500.

SWEIGART, A. L., FISHMAN, L. & WILLIS, J. H. (2006) A simple genetic incompatibility causes hybrid male sterility in *Mimulus*. *Genetics,* 172: 2465-2479.

TAKAMURA, T. & WATANABE, T. K. (1980) Further studies on the *Lethal hybrid rescue* (*Lhr*) gene of *Drosophila simulans*. *Japanese Journal of Genetics,* 55: 405-408.

TANG, S. & PRESGRAVES, D. C. (2009) Evolution of the *Drosophila* nuclear pore complex results in multiple hybrid incompatibilities. *Science,* 323: 779-782.

TING, C.-T., TSAUR, S.-C., SUN, S., BROWNE, W. E., CHEN, Y.-C., PATEL, N. H. & WU, C.-I. (2004) Gene duplication and speciation in *Drosophila*: evidence from the *Odysseus* locus. *Proceedings of the National Academy of Sciences,* 101: 12232-12235.

TING, C.-T., TSAUR, S.-C., WU, M.-L. & WU, C.-I. (1998) A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science,* 282: 1501-1504.

TOKARSKA-GUZIK, B. & DAJDOK, Z. (2007) NOBANIS - Invasive alien species fact sheet - *Mimulus guttatus*. *Online database of the North European and Baltic Network on Invasive Alien Species - NOBANIS.*

TURELLI, M. & ORR, H. A. (1995) The dominance theory of Haldane's rule. *Genetics,* 140, 389-402.

URADE, R. (2007) Cellular response to unfolded proteins in the endoplasmic reticulum of plants. *FEBS,* 274: 1152-1171.

VAN DER BIEZEN, E. A. & JONES, J. D. G. (1998) Plant disease-resistance proteins and the gene-for-gene concept. *Trends in Biochemical Sciences,* 23: 454-456.

VAN DER HOORN, R. A. L., WULFF, B. B. H., RIVAS, S., DURRANT, M. C., VAN DER PLOEG, A., DE WIT, P. J. G. M. & JONES, J. D. G. (2005) Structure-function analysis of Cf-9, a receptor-like protein with extracytoplasmic leucine-rich repeats. *The Plant Cell,* 17: 1000-1015.

VICKERY, R. K. (1964) Barriers to gene exchange between members of the *Mimulus guttatus* complex (Scrophulariaceae). *Evolution,* 18: 52-69.

VICKERY, R. K. (1978) Case studies in the evolution of complexes in *Mimulus*. *Evolutionary Biology,* 11: 405-507.

VICKERY, R. K. (1995) Speciation by aneuploidy and ploploidy in *Mimulus* (Scrophulariaceae). *Great Basin Naturalist*, 55(2): 174-176.

WATANABE, T. K. & YAMADA, M.-A. (1977) Effects of *SR* factor and *da* gene on the viability of the hybrid between *Drosophila melanogaster* and *D. Simulans*. *Japanese Journal of Genetics,* 52: 9-14.

WATANABE, T. K. (1979) A gene that rescues the lethal hybrids between *Drosophila melanogaster* and *D. Simulans*. *Japanese Journal of Genetics,* 54: 325-331.

WIDMER, A., LEXER, C. & COZZOLINO, S. (2009) Evolution of reproductive isolation in plants. *Heredity,* 102: 31-38.

WIEBE, G. A. (1934) Complementary factors in barley giving a lethal progeny. *The Journal of Heredity,* 25: 273-274.

WITTBRODT, J. ADAM, D., MALITSCHEK, B., MAUELER, W., RAULF, F., TELLING, A. ROBERTSON, S.M., SCHARTL, M. (1989) Novel putative receptor tyrosine kinase encoded by the melanoma-inducing *Tu* locus in *Xiphophorus*. *Nature,* 341: 415-421.

WU, C. A., LOWRY, D. B., COOLEY, A. M., WRIGHT, K. M., LEE, Y. W. & WILLIS, J. H. (2007) *Mimulus* is an emerging model system for the integration of ecological and genomic studies. *Heredity, Advance Online Publication June 6th 2007 (also publsihed in 2008, 100:220-230).*

WU, C. I. & TING, C. T. (2004) Genes and speciation. *Nature Reviews Genetics,* 5: 114-122.

WU, C. I. (2001a) The genic view of the process of speciation. *Journal of Evolutionary Biology,* 14: 851-865.

WU, C. I. (2001b) Genes and speciation – Reply. *Journal of Evolutionary Biology,* 14: 889-891.

WULFF, B. B. H., KRUIJT, M., COLLINS, P. L., THOMAS, C. M., LUDWIG, A. A., DE WIT, P. J. G. M. & JONES, J. D. G. (2004) Gene shuffling-generated and natural variants of the tomato resistance gene *Cf-9* exhibit different auto-necrosis-inducing activities in *Nicotiana* species. *The Plant Journal,* 40: 942-956.

YAMADA, T. & MARUBASHI, W. (2003) Overproduced ethylene causes programmed cell death leading to temperature-sensitive lethality in hybrid seedlings from the cross *Nicotiana suaveolens* x *N. tabacum*. *Planta,* 217: 690-698.

YAMADA, T., MARUBASHI, W. & NIWA, M. (1999) Detection of four lethality types in interspecific crosses among *Nicotiana* species through the use of three rescue methods for lethality. *Japanese Society of Breeding,* 49: 203-210.

YAMAMOTO, E., TAKASHI, T., MORINAKA, Y., LIN, S., KITANO, H., MATSUOKA, M. & ASHIKARI, M. (2007) Interaction of two recessive genes, *hbd2* and *hbd3*, induces hybrid breakdown in rice. *Theoretical and Applied Genetics,* 115: 187-194

YAMAMOTO, E., TAKASHI, T., MORINAKA, Y., LIN, S., WU, J., MATSUMOTO, T., KITANO, H., MATSUOKA, M. & ASHIKARI, M. (2010) Gain of deleterious function causes an autoimmune response and Bateson-Dobzhansky-Muller incompatibility in rice. *Molecular Genetics and Genomics,* 283: 305-315.

ZHANG, S. & LIU, Y. (2001) Activation of salicylic acid-induced protein kinase, a mitogen-activated protein kinase, induces defense responses in tobacco. *The Plant Cell,* 13: 1877-1889.

ZHU, Y., QIAN, W. & HUA, J. (2010) Temperature modulates plant defence responses through NB-LRR proteins. *PLoS Pathology,* 6(4): e1000844.

ZIPPERLEN, P., FRASER, A. G., KAMATH, R. S., MARTINEZ-CAMPOS, M. & AHRINGER, J. (2001) Roles for 147 embryonic lethal genes on *C. elegans* chromosome 1 identified by RNA interference and video microscopy. *EMBO,* 20: 3984-3992.

Websites

http://www.mimulusevolution.org
http://www.phytozome.net/
http://www.arabidopsis.org/
http://pfam.sanger.ac.uk/
http://www.genome.jp/kegg/
http://www.pantherdb.org/
http://frodo.wi.mit.edu/primer3/
http://imed.med.ucm.es/cgi-bin/emboss.pl?_action=input&_app=revseq
http://www.ebi.ac.uk/Tools/emboss/align/index.html
http://blast.ncbi.nlm.nih.gov/Blast.cgi
http://www.ncbi.nlm.nih.gov/projects/gorf/
www.genes.mit.edu/GENSCAN.html
http://faculty.vassar.edu/lowry/fisher2x3.html
http://www.arabidopsis.org/