This electronic thesis or dissertation has been downloaded from Explore Bristol Research, http://research-information.bristol.ac.uk

Author:
Hanley, Steven J

Title:
Genetic mapping of important agronomic traits in biomass willow

General rights
Access to the thesis is subject to the Creative Commons Attribution - NonCommercial-No Derivatives 4.0 International Public License. A copy of this may be found at https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode This license sets out your rights and the restrictions that apply to your access to the thesis so it is important you read this before proceeding.

Take down policy
Some pages of this thesis may have been removed for copyright restrictions prior to having it been deposited in Explore Bristol Research. However, if you have discovered material within the thesis that you consider to be unlawful e.g. breaches of copyright (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please contact collections-metadata@bristol.ac.uk and include the following information in your message:

• Your contact details
• Bibliographic details for the item, including a URL
• An outline nature of the complaint

Your claim will be investigated and, where appropriate, the item in question will be removed from public view as soon as possible.

Genetic Mapping of Important Agronomic Traits in Biomass Willow

Steven J. Hanley

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of PhD in the Faculty of Science.
Department of Agricultural Sciences
July 2003

67,890 words

Abstract

Increased energy production from renewable resources is currently a high priority in Europe as a result of dwindling fossil fuel reserves and the need to produce energy in ways that are non-polluting. The generation of energy from biomass sources can contribute towards a solution. Several willow species have been highlighted as particularly suitable for this end-use when grown as Short Rotation Coppice (SRC). Although efforts to breed high-yielding varieties for this purpose have resulted in significant improvements in recent years (Larsson, 2001; Lindegaard *et al.*, 2001), the scope for further progress remains great. However, current breeding efforts are hampered by inefficient selection regimes and the limited knowledge available on the genetic basis of important agronomic traits. The main objective of this thesis was to use molecular markers in genetic mapping strategies to address both of these issues.

Two genetic linkage maps, based on two different full-sib biomass willow mapping populations, were generated, largely based on microsatellite and AFLP markers. Both maps were produced according to a double pseudo-testcross strategy in which separate parental maps were constructed before integration to form a consensus map (Grattapaglia and Sederoff, 1994). The first map was derived from a pre-existing population (K3), comprised of only 66 progeny. A second, much larger, full-sib population, K8, was established to underpin subsequent trait analyses and QTL mapping. This population comprised 947 full-sib progeny and was planted in a field trial at Long Ashton Research Station. This trial was used for assessments of traits of agronomic importance, including total yield, several components of yield (e.g. height, diameter, etc.), resistance to *Melampsora* rust diseases and resistance to willow beetle herbivory. Laboratory-based assessments of disease and pest resistance were also performed. All resulting trait data were used in conjunction with genotype and linkage information to map genomic regions influencing these traits via QTL analyses. Putative QTL for all target traits were identified, with indications of robustness provided by the identification of corresponding QTL positions for correlated traits and by the detection of several QTL that were consistent over different assessment years and environments.

These data provide the most comprehensive QTL mapping study of traits of importance in biomass willow performed to date and represent a valuable resource for future willow research and molecular breeding of improved biomass varieties.

Acknowledgments

Sincere thanks to my supervisor Angela Karp, for giving me the opportunity to take on this project and for providing much valued support and advice throughout its duration, particularly in the final year which coincided with a number of logistical hurdles which arose from the closure of LARS.

Thanks also to Jackie Barker, for guidance, enthusiasm and expert teaching in the initial stages of the project, and for the constant encouragement and support throughout my entire time at LARS.

I would also like to acknowledge the contributions made by all members of the 'willow team', whose advice and discussion was invaluable - Tom Hunter, Ming Pei, Lori Peacock, Carmen Ruiz, Kevin Lindegaard and Rod Parfitt.

Many thanks to the LARS experimental husbandry staff, for assistance and expertise relating to many aspects of the fieldwork, particularly the 'somewhat repetitive' trait assessments performed at both harvests.

Statistical support and guidance was kindly provided by Stephen Powers, Robin Thompson, Gillian Arnold and Phil Brain. Particular thanks to Steve, for the rapid responses to my 'last-minute' questions and for finding the time to look at the data generated at such a late stage in the project.

Thanks also to Johan Van Ooijen for helpful discussion and advice.

Thanks also to everyone in Barker Lab 18 - Rachael, Sveta, Chloe, Carlos, Jenny for support and discussion. I am particularly grateful to Sarah for technical support and help with field assessments (and for improving my knowledge of British wildlife!).

I am grateful for the support provided by both my friends and family – hopefully things will get back to normal now. Many thanks to James for all the help with IT needs at home. Above all, I would like to thank Louise for being patient, understanding and keeping me motivated.

This work was funded under the Department of Trade and Industry's Renewable Energy Programme (managed by ETSU), UK (Agreement No. B/W6/00599/00/00). Thanks to Mr. Nick Beale (ETSU, UK) for guidance and support. LARS received grant-aided support from the BBSRC.

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the Regulations of the University of Bristol. The work is original except where indicated by special reference in the text and no part of the dissertations has been submitted for any other degree.

Any views expressed in the dissertation are those of the author and in no way represent those of the University of Bristol.

The dissertation has not been presented to any other university for examination either in the United Kingdom or overseas.

Signed: Date: 8.12.03

Contents

Chapter 1. Introduction 2

1.1 THE GENUS *SALIX* 4

1.2 SHORT ROTATION COPPICE (SRC) WILLOW AS A BIOMASS CROP 5

1.2.1 Energy production from SRC 7

1.2.2 Growing SRC willows 8

1.2.3 Threats to SRC yield 9
1.2.3.1 Diseases of SRC willow 9
1.2.3.1.1 Rust diseases 9
1.2.3.1.2 Disease management . 12
1.2.3.2 Invertebrate pests 12
1.2.3.2.1 Leaf-feeding willow beetles 12
1.2.3.2.2 Willow beetle management strategies 15
1.2.3.2.3 Other insect pests 16
1.2.3.3 Mammalian herbivores 16

1.2.4 Breeding of improved biomass willows 17
1.2.4.1 Historical overview 17
1.2.4.2 Breeding objectives and strategy 18
1.2.4.3 Breeding practicalities 19
1.2.4.4 Current breeding and selection schemes 20
1.2.4.5 Future breeding prospects 20

1.3 GENETIC MARKERS 21

1.3.1 Morphological markers 21

1.3.2 Molecular markers 21
1.3.2.1 Restriction Fragment Length Polymorphism (RFLP) 22
1.3.2.2 Arbitrary priming marker classes 23

1.3.2.3 Amplified Fragment Length Polymorphism (AFLP) 24
1.3.2.4 Sequence-based PCR markers 24
1.3.2.4.1 Microsatellite markers 25
1.3.2.4.2 Markers based on expressed sequences 28
1.3.2.4.3 Single Nucleotide Polymorphisms (SNPs) 29
1.3.2.5 Molecular marker summary 29

1.4 THE USE OF MOLECULAR MARKERS IN CROP IMPROVEMENT 30

1.4.1 Investigations into germplasm relationships 30
1.4.1.1 Conservation and assessment of genetic variation within gene pools 31
1.4.1.2 Exploitation of genetic variation in breeding programmes 32

1.4.2 Linkage mapping 32
1.4.2.1 The basic underlying principles of linkage mapping 33
1.4.2.2 Linkage mapping in practice 37
1.4.2.2.1 Mapping populations 37
1.4.2.2.2 Choice of molecular marker class 38
1.4.2.3 Applications of linkage maps in crop improvement 38

1.4.3 Genetic mapping of important agronomic traits 39
1.4.3.1 Analysis of quantitative traits using molecular markers: QTL analysis 41
1.4.3.1.1 QTL mapping methodologies 42
1.4.3.1.2 Applications of QTL analysis in crop improvement 44

1.4.4 Comparative mapping 45

1.4.5 Marker-assisted selection 47

1.4.6 Map-based cloning 50

1.4.7 The candidate gene approach 51

1.5 SPECIFIC AIMS AND OBJECTIVES OF THIS THESIS 54

1.6 CHAPTER SUMMARY 56

Chapter 2. General Materials and Methods 57

2.1 BIOLOGICAL MATERIALS 59

2.1.1 Plant material 59

2.1.2 Willow beetles 59

2.2 NON-BIOLOGICAL MATERIALS 59

2.2.1 Chemicals 59

2.2.2 Radiochemicals 59

2.2.3 Restriction and modifying enzymes 60

2.2.4 Oligonucleotides 60

2.3 STANDARD SOLUTIONS 60

2.3.1 Molecular biology solutions 60

2.3.2 Liquid bacterial growth medium 61

2.4 GENERAL MOLECULAR BIOLOGY METHODS 61

2.4.1 Isolation of total plant genomic DNA from leaf tissue 61

2.4.2 Agarose gel electrophoresis 62

2.4.3 Preparation of PCR products for direct sequencing 62

2.4.4 Automated DNA sequencing 63

2.4.5 Sequence analysis 63

2.5 MOLECULAR MARKER METHODS 63

2.5.1 Microsatellite methods 63

2.5.1.1 Characterisation of microsatellite library inserts 63

2.5.1.2 Isolation of bacterial plasmid DNA and sequencing of inserts 64
2.5.1.3 Microsatellite primer design 64
2.5.1.4 Microsatellite amplification and detection 65
2.5.1.4.1 Microsatellite amplification and detection : radioisotope method 65
2.5.1.4.2 Microsatellite amplification and detection: [F]dCTP method 66
2.5.1.4.3 Microsatellite amplification and detection: fluorescent primers 66

2.6.1 AFLP methods 66
2.6.1.1 AFLP Method 1 (based on Zabeau and Vos, 1993) 66
2.6.1.2 AFLP Method 2 (Vos *et al.*, 1995) 67

2.6.2 AFLP visualisation 68

2.7 TRAIT ASSESSMENTS 69

2.8 STATISTICAL METHODS 69

2.8.1 Linkage analysis 69

2.8.2 QTL analysis 70

2.8.3 Statistical analysis of trait data 70

Chapter 3. Construction of the K3 Linkage Map 71

3.1 INTRODUCTION 73

3.1.1 Genetic mapping in willow: general considerations 73

3.1.2 Linkage analysis in outbreeding species 73

3.2 MATERIALS AND METHODS 76

3.2.1 The K3 mapping population 76

3.2.2 DNA extraction 77

3.2.3 Development of additional AFLP markers 77

3.2.4 Scoring of AFLP markers 77

3.2.5 Development of additional microsatellite markers 77

3.2.6 Microsatellite primer testing and screening 78

3.2.7 Scoring microsatellite markers 79

3.2.8 Poplar microsatellite testing 79

3.2.9 Sequencing of poplar microsatellite products at locus PMGC223 80

3.2.10 ESTP markers 80

3.2.11 Map construction 81

3.3 RESULTS 82

3.3.1 AFLP markers 82

3.3.2 Microsatellite markers: development and screening 84

3.3.3 Poplar microsatellite primers 88

3.3.4 ESTP markers 89

3.3.5 Construction of the parental maps 89

3.3.6 Construction of the K3 consensus map 95

3.4 DISCUSSION 107

3.4.1 Choice of molecular markers 107

3.4.2 AFLP markers 108

3.4.3 Willow microsatellite markers 108

3.4.4 Poplar microsatellites 109

3.4.5 ESTP markers 110

3.4.6 Segregation analysis 111

3.4.7 The K3 parental maps 112

3.4.8 The K3 consensus map 112

3.5 CHAPTER SUMMARY 114

Chapter 4. Selection and Establishment of a Large Segregating Population as a Basis for QTL Mapping 115

4.1 INTRODUCTION 117

4.1.1 Selection of candidate populations 118

4.1.2 Increasing the number of K3 progeny 122

4.1.3 Experimental design 122

4.1.4 Pedigree authentification 125

4.2 MATERIALS AND METHODS 125

4.2.1 DNA extractions for pedigree authentification 125

4.2.2 AFLP analysis of candidate pedigrees 125

4.2.3 Pedigrees of the chosen crosses K1 and K8 126

4.2.4 Preparation of cuttings for planting the field trial 127

4.2.5 Planting the field trial 128

4.2.6 Field-based trait assessments 128

4.2.6.1 Field-based assessments of rust resistance 129

4.2.6.2 Destructive yield assessments 129

4.2.6.3 Stem diameter assessments 130

4.2.6.4 Stem height assessments 130

4.2.7 Laboratory-based trait assessments 130

4.2.7.1 Laboratory-based assessment of rust resistance 131

4.2.7.2 Laboratory-based beetle feeding experiments 132

4.3 RESULTS 135

4.3.1 AFLP analysis for authentification of parents used in candidate pedigrees 135

4.3.2 Establishment of the field trial 137

4.3.3 Trait assessments 138

4.3.4 Laboratory-based trait assessments 142

4.3.4.1 Laboratory-based assessment of rust resistance 142

4.3.4.2 Laboratory-based beetle feeding experiments 144

4.4 DISCUSSION 145

4.4.1 AFLP analysis of candidate pedigrees 145

4.4.2 Field-based trait assessments 146

4.4.2.1 Field-based assessment of rust resistance 146

4.4.2.2 Assessments relating to yield 148

4.4.3 Laboratory-based assessments 148

4.4.3.1 Laboratory-based assessment of rust resistance 149

4.4.3.2 Laboratory-based beetle feeding experiments 149

4.5 CHAPTER SUMMARY 151

Chapter 5. Construction of the K8 Linkage Map 152

5.1 INTRODUCTION 154

5.1.1 General strategy 154

5.1.2 Pedigree authentification 155

5.2 MATERIALS AND METHODS 157

5.2.1 Plant material 157

5.2.2 High-throughput DNA extraction from K8 progeny 157

5.2.3 Identification of the correct K8 parents 157

5.2.4 Development of additional SSR markers 158

5.2.5 Testing and screening of SSR primer sets 158

5.2.5.1 Screening microsatellites for detection by autoradiography 159

5.2.5.2 Screening microsatellites for fluorescent detection 159

5.2.6 Development of microsatellite multiplexes 160

5.2.7 Screening microsatellite multiplexes 161

5.2.8 Scoring microsatellite markers 163

5.2.9 Bulked segregant analysis 163

5.2.10 AFLP screening 165

5.2.11 Map construction 165

5.3 RESULTS 166

5.3.1 Identification of the K8 parents 166

5.3.2 Development of additional microsatellite markers 168

5.3.3 Testing for microsatellite polymorphisms in K8 169

5.3.4 Development of microsatellite mutiplexes 169

5.3.5 Microsatellite screening of the K8 population 172

5.3.6 Bulked segregant analysis 174

5.3.7 AFLP markers 177

5.3.8 Construction of the parental maps 178

5.3.9 The K8 consensus map 182

5.3.10 Mapping AFLP markers putatively linked to microsatellite SB226 183

5.3.11 Identification of homeologous K3 and K8 linkage groups 189

5.4 DISCUSSION 194

5.4.1 Identification of the K8 parents 194

5.4.2 Microsatellite markers 195

5.4.2.1 Microsatellite marker development 195

5.4.2.2 Microsatellite screening 196

5.4.3 AFLP markers 197

5.4.4 Bulked segregant analysis 198

5.4.5 The K8 parental maps 199

5.4.6 The K8 consensus map 200

5.4.7 Comparisons with the K3 map 201

5.5 CHAPTER SUMMARY 202

Chapter 6. Trait Analysis and QTL Mapping 204

6.1 INTRODUCTION 206

6.1.1 QTL analysis in outbreeding species 206

6.2 MATERIAL AND METHODS 206

6.2.1 Additional trait assessments 206

6.2.1.1 Field-based assessments of rust resistance 207

6.2.1.2 Non-destructive yield assessments 207

6.2.1.3 Destructive yield assessments 207

6.2.1.4 Laboratory-based beetle feeding experiment 207

6.2.2 Examination of spatial variability and adjustment of K8 trait values 208

6.2.3 Phenotypic correlations between characters 211

6.2.4 QTL analysis 211

6.3 RESULTS **214**

6.3.1 Additional trait assessments 214

6.3.1.1 Field-based trait assessments 214

6.3.1.2 Laboratory-based beetle feeding assay 217

6.3.2 Adjustment of K8 trait values for spatial heterogeneity 218

6.3.3 Correlations between characters 219

6.3.3.1 Correlations between growth traits in the establishment year 222

6.3.3.2 Correlations between growth traits in the second year (2001-2002) 223

6.3.3.3 Correlations between growth traits in the third year (2002-2003) 223

6.3.3.4 Correlations between traits over three years 224

6.3.4 QTL analysis I. K8 225

6.3.4.1 The K8 paternal data set 226

6.3.4.2 The K8 maternal data set 229

6.3.4.3 The K8 consensus data set 231

6.3.5 QTL analysis II. K3 239

6.3.5.1 The K3 paternal data set 239

6.3.5.2 The K3 maternal data set 242

6.3.5.4 The K3 consensus data set 246

6.4 DISCUSSION **252**

6.4.1 Fundamental characteristics of the traits assessed 252

6.4.1.1 Growth traits 252

6.4.1.2 Rust resistance 253

6.4.2 Correlations between characters in K8 254

6.4.3 QTL analysis: Approach 256

6.4.3.1 QTL analysis : K8 258

6.4.3.1.1 General considerations 258

6.4.3.1.2 QTL detection : K8 259

6.4.3.2 QTL analysis : K3 264

6.4.3.2.1 General considerations 264

6.4.3.2.2 QTL detection : K3 264

6.5 CHAPTER SUMMARY 267

Chapter 7. General Discussion and Future Perspectives 269

References 279

List of Figures

1.1	The three main technologies used to generate heat and/or power from SRC	7
1.2	A SRC willow plantation	9
1.3	The life cycles of *Melampsora epitea* and *M. capraearum*	10
1.4	A willow leaf infected with rust	11
1.5	The three main willow beetle species that cause damage to SRC plantations – *Phratora vulgatissima* (A), *Phratora vitellinae* (B) and *Galerucella lineola* (C)	13
1.6	Leaf damage caused by *Phratora vulgatissima* – the blue willow beetle	13
1.7	Sequence electropherogram of a microsatellite locus showing a perfect di-nucleotide $[GT]_{15}$ repeat motif flanked by unique sequence to which primers can be designed for interrogation of size polymorphisms by PCR	26
1.8	Schematic representation of recombination by crossover events during meiosis	34
1.9	Diagram to show how a maximum of 50 % recombinants can result from a single crossover event involving two chromatids at the four-strand stage of meiosis. Maximum recombination frequencies can only be obtained when two loci are located at opposite ends of a chromosome	35
1.10	A simple representation of a hypothetical genetic linkage map based on three loci A, B and C (for explanation, see text)	35
1.11	Diagram to show how double or multiple crossover events can prevent the detection of recombination between linked loci	35
1.12	Conceptual models for quantitative inheritance. Graphs represent the F2 progeny derived from the selfing of a F1 generation derived from a cross between homozygous parents	40
1.13	A LOD profile identifying a theoretical QTL by interval mapping. The most likely position of the QTL corresponds to the highest peak of the LOD profile and is located at the 12 cM position. The dotted horizontal line represents the LOD significance threshold, below which, the LOD score is statistically insignificant	43

3.1 A LOD profile identifying a theoretical QTL by interval mapping. The most likely position of the QTL corresponds to the highest peak of the LOD profile and is located at the 12 cM position. The dotted horizontal line represents the LOD significance threshold, below which, the LOD score is statistically insignificant 76

3.2 A section of an AFLP mapping gel generated with primer combination MACA/PAC. Profiles for the K3 mapping population, along with those of the grandparents and parents are shown. The three possible marker segregation types are also indicated 82

3.3 Autoradiogram showing segregation of three alleles at polymorphic microsatellite locus SB337 85

3.4 An example of a [F]dCTP mapping gel showing fluoresecent detection of microsatellite alleles at two loci SB522 and SB565, both of segregation type abxab 86

3.5 Alignment of edited sequence data for microsatellite locus PMGC223 in *S. viminalis* var. Astrid", *S. viminalis*, "Germany" and poplar hybrid, "Ghoy". Shaded regions highlight positions at which all three sequences are equivalent. 'x' denotes that sequence could not be determined. '-' denotes base not present. 'N' indicates an ambiguous base 88

3.6 The K3 paternal linkage map 91

3.7 The K3 maternal linkage map 93

3.8 The K3 consensus linkage map 98

4.1 The final field trial plan comprising the K1, K8 and K3 mapping populations. The plots comprising six replicates of each of the K1, K8 and K3 genotypes are labelled 1001-1946, 8001-8947 and 3001-3209, respectively. Shaded boxes labelled Jorr indicate plots comprising *S. viminalis* var. Jorr for use as a control of spatial heterogeneity across the site. Additional shaded areas represent the planting of Jorr as buffer or tramline rows. Figure 4.1a shows the entire field plan, Figure 4.1b shows a magnified view of a section of the field trial and Figure 4.1c illustrates the planting arrangement within and between plots (not to scale). 124

4.2 Expected pedigrees of the K1 and K8 crosses 126

4.3 The K1 and K8 mapping populations growing in the nursery at LARS 127

4.4 Planting the field trial 128

4.5 Schematic representation of leaf disc layout for Petri-dish 1, population K1 133

4.6 An AFLP autoradiograph showing profiles generated for all candidate parents and grandparents of the mapping population. Pedigree information is provided in Table 4.1 136

4.7 The field trial at LARS in the second year of growth 138

4.8 Field- and nursery-based trait value distribution plots 139

4.9 An example of a single repli-plate showing variation in the number of observed uredinia (pathotype LET5) for five individuals of the K1 population 142

4.10 Distribution plots for mean pustule count data for the K1, K8 and K3 populations, following inoculation with two rust pathotypes (LET1 and LET5) 143

4.11 Variation in the amount of leaf area remaining for the ten leaf discs included in a single Petri-dish (Dish 2; Appendix I) of the K1 beetle feeding experiment 144

4.12 Distributions of beetle feeding test results for populations K1, K8 and K3 144

5.1 Microsatellite-based exclusion experiment to identify possible parents of the K8 mapping population. Analysis of five loci showed K477-S3 and K477-R13 to be the only possible parents of the mapping population 166

5.2 Correct pedigree information for the K8 mapping population 168

5.3 An example of successful genotype detection using a PCR- and loading-based multiplexing strategy 171

5.4 An autoradiograph of a section of a polyacrylamide gel showing two loadings of PCR products for microsatellite SB288. Three alleles (a, b and c) were detected at this locus 172

5.5 Bulked segregant analysis AFLP profiles identifying two markers putatively linked to microsatellite locus SB226 175

5.6 The K8 paternal linkage map 180

5.7 The K8 maternal linkage map 181

5.8 The K8 consensus map 184

5.9 Alignment of putatively homeologous K3 and K8 linkage groups. Microsatellite markers used to anchor corresponding linkage groups are shown in red type. K3 linkage groups are indicated by roman numerals unless otherwise stated. K8 linkage groups are indicated by standard alphabetic names 191

5.10 Alignment of K3 linkage group III and K8 linkage group B based on microsatellite loci SB276 and SB419b (=SB1075) 193

6.1 Field-based trait distribution plots : K8 214

6.2 Field-based trait distribution plots : K3 216

6.3 Distribution of logit transformed means of percentage leaf area remaining for the 2001 beetle feeding assessment 217

6.4 Variation in the amount of leaf area remaining for the ten leaf discs included in a single Petri-dish (Dish 2; Appendix I) of the K1 beetle feeding experiment 218

6.5 Scatter plot of MxHt-01 vs. FW-01 showing a non-linear relationship between maximum stem height and fresh weight (2001 data). 222

6.6 Scatter plots of MxHt-03 vs. FW-03 and MxDia-03 vs. FW-03 showing non-linear relationships between maximum stem height and diameter with fresh weight (2003) 223

6.7 Results of Kruskal-Wallis analysis for rust traits on consensus linkage group I. The significance of each marker-trait association is indicated by asterisks (*: P < 0.1, **: P < 0.05, ***: P < 0.01, ****: P < 0.005, *****: P < 0.001, ******: P < 0.0005, *******, P < 0.0001). Marker nomenclature is as described in Figure 3.6 236

6.8 Results of Kruskal-Wallis analysis for rust traits on K8 consensus linkage group B. The significance of each marker-trait association is indicated by asterisks (*: P < 0.1, **: P < 0.05, ***: P < 0.01, ****: P < 0.005, *****: P < 0.001, ******: P < 0.0005, *******, P < 0.0001). Marker nomenclature is as described in Figure 3.6 237

6.9 Results of interval mapping in population K8. Parental and consensus 238

linkage groups harbouring putative QTL are shown

6.10 Results of Kruskal-Wallis analysis for rust traits on K3 consensus linkage group III. The significance of each marker-trait association is indicated by asterisks (*: P < 0.1, **: P < 0.05, ***: P < 0.01, ****: P < 0.005, *****: P < 0.001, ******: P < 0.0005, *******, P < 0. 0001). Marker nomenclature is as described in Chapter 3, Figure 3.8 246

6.11 Results of interval mapping in population K3. Parental and consensus linkage groups harbouring putative QTL are shown 251

List of Tables

1.1	The three main subgenera of the genus *Salix*	4
1.2	Fundamental objectives of British and Swedish biomass willow breeding programmes	18
1.3	Approximate relationships between map distance and physical distance in different organisms	37
2.1	Marker segregation codes for use in JoinMap® 3.0	69
3.1	Details of the nineteen poplar microsatellite primers tested in willow	79
3.2	Details of Populus EST sequences used in ESTP development. Original EST names (Sterky et al., 1998) are shown above GenBank accession numbers (in italics)	81
3.3	Characteristics of polymorphic AFLP markers generated from 48 primer combinations. Restriction enzymes/adaptors/primers used in AFLP production are indicated for each combination (E = *Eco*RI, M = *Mse*I and P = *Pst*I). Additional characters indicate selective nucleotides. Primer combinations previously screened in the earlier project by Dr J.H.A Barker are shown in italics	83
3.4	Segregation characteristics of microsatellite markers included in the final K3 mapping data set. Markers developed and screened during the EU-AIR project are shown in italics	86
3.5	Figure 3.5. Alignment of edited sequence data for microsatellite locus PMGC223 in *S. viminalis* var. Astrid", *S. viminalis*, "Germany" and poplar hybrid, "Ghoy". Shaded regions highlight positions at which all three sequences are equivalent. 'x' denotes that sequence could not be determined. '-' denotes base not present. 'N' indicates an ambiguous base.	88
4.1	Details of the seven willow crosses originally proposed as candidate mapping populations	120
4.2	K477 population field-based rust scores recorded in 1996 and 1997 in Sweden. Scores were assigned on a scale ranging from 0 – 6 where 0 = no rust found: 6 = severe rust infection (Åhman, 1997). Mean values derived from assessment of three clonal replicates per K477 individual are provided. Parental (SW880435 var. Astrid and SW930984) and grandparental (SW880435 var. Astrid and SW910006 var. Björn) scores are also shown	121

4.3 Individuals selected from the K1, K8 and K3 mapping populations for inclusion in rust inoculation tests 131

4.4 K1, K8 and K3 individuals included in laboratory-based beetle feeding assays 133

4.5 Individuals of the K1 and K8 populations that failed to yield sufficient material for the production of six cuttings 137

4.6 Summary of trait data recorded during the field trial establishment year for populations K1 and K8. Data recorded from reference plots containing S. viminalis var. Jorr from within the K1 population, within the K8 population and within both populations, are shown. Data recorded from four replicated plots comprising the K1 and K8 parents and grandparents are also included. For all traits, mean, minimum and maximum scores are shown 141

4.7 Summary of rust inoculation experiment data. Values refer to mean pustule counts calculated from five replicates per genotype 143

4.8 Summary of results from beetle feeding assays for populations K1, K8 and K3. Values indicate the mean amount of leaf area remaining (mm^2) calculated from the ten replicates per genotype 145

5.1 Microsatellite multiplexes used for K8 progeny screening. All figures represent µl volumes used in PCR reactions. Colours refer to fluorescent labels 161

5.2 K8 individuals included in DNA pools used for bulked segregant analysis aimed at identifying AFLP markers linked to microsatellite locus SB226 163

5.3 AFLP primer combinations used to screen DNA pools in bulked segregant analysis where '✓' =tested; 'n/t' = not tested 164

5.4 Details of microsatellite loci included in the K8 mapping data set 173

5.5 Details of primer combinations yielding putatively informative markers from BSA. The number of markers that segregated only between the SB226-targeted bulks, only between the rust resistant/susceptible bulks and between both the SB226-targeted and rust resistant/susceptible bulks are shown. Figures in brackets indicate a marker that could not be conclusively scored 176

5.6 Characteristics of polymorphic AFLP markers generated from seven primer combinations. Restriction enzymes/adapters/primers used in AFLP production are indicated for each combination (E = *Eco*RI, M = *Mse*I and P = *Pst*I). Additional characters indicate selective nucleotides 177

6.1 Summary of traits included in QTL analyses. Details of the timing of assessments and the QTL methods used for analysis (KW = Kruskal-Wallis, IM = interval mapping) for each trait are provided. Unless specified, traits were assessed for both K8 and K3. Codes assigned for each trait are indicated 212

6.2 Correlation data for all traits assessed in the K8 field trial. Strong correlations (r > 0.5) are shown in bold type 220

6.3 Associated probabilities for correlations. Non-significant correlations (P > 0.0003; Section 6.2.3) are shown in red 221

6.4 Summarised results of QTL analyses of the K8 paternal (K477-S3) data set for yield-related traits. For explanation see text 228

6.5 Summary of putative QTL identified for pest and disease traits in the K8 male parent (K477-R13) 229

6.6 Summarised results of QTL analyses of the K8 maternal (K477-S3) data set for yield-related traits. For explanation see text 230

6.7 Summary of putative QTL identified for pest and disease traits in the K8 female parent (K477-S3) 231

6.8 Summarised results of QTL analyses of the K8 consensus data set for yield-related traits. For explanation see text 233

6.9 Summarised results of QTL analyses of the K8 consensus data set for disease and pest traits. For explanation see text. 235

6.10 Summarised results of QTL analyses based on growth traits for the K3 male parent (SW980082 var. Orm). See text for explanation 241

6.11 Summarised results of QTL analyses based on pest and disease traits in the K3 male parent (SW980082 var. Orm). See text for explanation 242

6.12 Summarised results of QTL analyses based on growth-related traits in the K3 female parent (SW980084). For explanation see text. 244

6.13 Summarised results of QTL analyses based on pest and disease traits in the K3 female parent (SW980084). For explanation see text 245

6.14 Summarised results of QTL analyses based on growth-related traits and the K3 consensus data (SW980084). For explanation see text 248

6.15 Summarised results of QTL analyses based on pest and disease traits and the K3 consensus data. For explanation see text 250

Abbreviations and Acronyms

Units of length

m	metre	
cm	centimetre	(10^{-2} m)
mm	millimetre	(10^{-3} m)

Units of volume

l	litre	
ml	millilitre	(10^{-3} l)
µl	microlitre	(10^{-6} l)

Units of mass

kg	kilogram	(10^{3} g)
g	gram	
mg	milligram	(10^{-3} g)
µg	microgram	(10^{-6} g)
ng	nanogram	(10^{-9} g)
pg	picogram	(10^{-12} g)

Number of molecules

mmol	millimole	(10^{-3} mol)
µmol	micromole	(10^{-6} mol)
nmol	nanomole	(10^{-9} mol)
pmol	picomole	(10^{-12} mol)

Units of concentration

M	molar	
mM	millimolar	(10^{-3} M)
µM	micromolar	(10^{-6} M)

g / l grams per litre

μg / μl micrograms per microlitre

ng / μl nanograms per microlitre

% percent (where 1 g (1ml) in 100 ml is 1%)

Units of time

h hour

min minute

s second

Other units

°C degrees centigrade

rpm revolutions per minute

μCi microcurie

V volts

W watts

mA milliampere

U enzyme units

bp nucleotide base pair

kb kilobase pair (10^3 bp)

cM centimorgan

M morgan

Nucleic acid abbreviations

A adenine

C cytosine

G guanine

T thymine

Other abbreviations and acronyms

1 x	one times concentration
21-mer	oligonucleotide of 21 base pair length
3'	3 prime
5'	5 prime
ABI	Applied Biosystems
AB-QTL	advanced-backcrosss QTL analysis
AFLP	amplified fragment length polymorphism
ANOVA	analysis of variance
AP-PCR	arbitrary-primed polymerase chain reaction
ATP	adenosine-5'-triphosphate
BAC	bacterial artificial chromosome
BBSRC	Biotechnology and Biological Sciences Research Council, UK
BSA	bulked segregant analysis
cDNA	complementary DNA
©	copyright
CP	cross-pollinator
DAF	DNA amplification fingerprinting
DH	double haploid
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleoside-5'-triphosphate
EDTA	ethylene diaminetetra-acetic acid
EST	expressed sequence tag
ESTP	expressed sequence tag polymorphism
EU	European Union
EWBP	European Willow Breeding Programme
F_1	first filial generation
F_2	second filial generation
FAM	6-carboxyfluorescein
h/d	hours per day
HCl	hydrochloric acid
IM	interval mapping
KCl	potassium chloride

KW	Kruskal-Wallis analysis
LARS	Long Ashton Research Station
LARS	Long Asthon Research Station
LB	Luria Bertani
LET	*larici-epitea typica*
LG	linkage group
LOD	logarithm of odds
LR	*larici-retusae*
LSD	least significant difference
MAS	marker-assisted selection
$MgCl_2$	magnesium chloride
mtoe	miliion tonnes of oil equivalent
NWC	National Willow Collection (held at LARS)
OPA	One-Phor-All buffer
P	probability
PAL	phenylalanine ammonia-lyase
PCR	polymerase chain reaction
pH	the relative hydrogen ion concentration in a solution (acidity/alkalinity)
PNK	poly nucleotide kinase
QTL	quantitative trait locus/loci
®	registered trademark
RAPD	random amplified polymorphism
REML	residual maximum likelihood
RFLP	restriction fragment length polymorphism
RGA	resistance gene analogue
RILs	recombinant inbred lines
RNA	ribonucleic acid
RT-PCR	reverse transcription PCR
SCN	soybean cyst nematode
SDW	sterile distilled water
SNP	single nucleotide polymorphism
SRC	short rotation coppice
SSCP	single stranded conformational polymorphism
SSLP	simple sequence length polymorphism

SSR	simple sequence repeat
Taq	*Thermus aquaticus*
TBE	Tris-Borate-EDTA buffer
TE	Tris-EDTA buffer
TEMED	N,N,N',N'-tetramethylethylenediamene
™	trademark
T_m	melting tempertature
Tris	2-amino-2-hydroxymethylpropan-1,3-diol
UK	United Kingdom
US	United States
YAC	yeast artificial chromosome

Note: ROX, VIC, NED, LIZ: full product information not yet available from manufacturer (Applied Biosystems).

Publications

Hanley S, Barker JHA, Van Ooijen JW, Aldam C, Harris S, Åhman I, Larsson S and Karp A. A genetic linkage map of willow (*Salix viminalis*) based on AFLP and microsatellite markers. *Theoretical and Applied Genetics* **105**, 1087-1096.

Danquah EY, Hanley SJ, Brookes RC, Aldam C and Karp A. 2002. Isolation and characterization of microsatellites in Echinochloa (L.) Beauv. spp. *Molecular Ecology Notes* **2**, 54-56.

Conference abstracts

Hanley SJ, Barker JHA, Aldam C, Lindegaard KN, Pei M, Hunter T and Karp A. 2000. Improving breeding efficiency of biomass willows using molecular marker technology. Plant and Animal Genome Meeting VIII, San Diego, 9th- 12th January 2000. (oral presentation)

Hanley SJ, Barker JHA and Karp A. 2000. Improving willow breeding efficiency. DTI SRC seminar series: Breeding, pest and disease control, Blagdon, October 4th 2000. (oral presentation)

Hanley SJ, Harris S and Karp A. 2002. Genetic mapping of important agronomic traits in biomass willow. International Poplar Symposium, Uppsala, Sweden, 26th – 29th August 2002. (poster presentation)

Hanley SJ, Harris S and Karp A. 2003. Genetic mapping of important agronomic traits in biomass willows. Plant and Animal Genome Meeting XI, San Diego, 11th – 15th January, 2003. (oral presentation)

Hanley SJ, Harris S and Karp A. 2003. Genetic mapping of important traits in biomass willows. Tree genomics workshop, SEB meeting, Southampton, 1st April 2003. (oral presentation)

PAGE MISSING IN ORIGINAL

1. Introduction

1.1 THE GENUS *SALIX* 4

1.2 SHORT ROTATION COPPICE (SRC) WILLOW AS A BIOMASS CROP 5

1.2.1 Energy production from SRC 7

1.2.2 Growing SRC willows 8

1.2.3 Threats to SRC yield 9
1.2.3.1 Diseases of SRC willow 9
1.2.3.1.1 Rust diseases 9
1.2.3.1.2 Disease management 12
1.2.3.2 Invertebrate pests 12
1.2.3.2.1 Leaf-feeding willow beetles 12
1.2.3.2.2 Willow beetle management strategies 15
1.2.3.2.3 Other insect pests 16
1.2.3.3 Mammalian herbivores 16

1.2.4 Breeding of improved biomass willows 17
1.2.4.1 Historical overview 17
1.2.4.2 Breeding objectives and strategy 18
1.2.4.3 Breeding practicalities 19
1.2.4.4 Current breeding and selection schemes 20
1.2.4.5 Future breeding prospects 20

1.3 GENETIC MARKERS 21

1.3.1 Morphological markers 21

1.3.2 Molecular markers 21
1.3.2.1 Restriction Fragment Length Polymorphism (RFLP) 22
1.3.2.2 Arbitrary priming marker classes 23
1.3.2.3 Amplified Fragment Length Polymorphism (AFLP) 24

1.3.2.4 Sequence-based PCR markers 24

1.3.2.4.1 Microsatellite markers 25

1.3.2.4.2 Markers based on expressed sequences 28

1.3.2.4.3 Single Nucleotide Polymorphisms (SNPs) 29

1.3.2.5 Molecular marker summary 29

1.4 THE USE OF MOLECULAR MARKERS IN CROP IMPROVEMENT 30

1.4.1 Investigations into germplasm relationships 30

1.4.1.1 Conservation and assessment of genetic variation within gene pools 31

1.4.1.2 Exploitation of genetic variation in breeding programmes 32

1.4.2 Linkage mapping 32

1.4.2.1 The basic underlying principles of linkage mapping 33

1.4.2.2 Linkage mapping in practice 37

1.4.2.2.1 Mapping populations 37

1.4.2.2.2 Choice of molecular marker class 38

1.4.2.3 Applications of linkage maps in crop improvement 38

1.4.3 Genetic mapping of important agronomic traits 39

1.4.3.1 Analysis of quantitative traits using molecular markers: QTL analysis 41

1.4.3.1.1 QTL mapping methodologies 42

1.4.3.1.2 Applications of QTL analysis in crop improvement 44

1.4.4 Comparative mapping 45

1.4.5 Marker-assisted selection 47

1.4.6 Map-based cloning 50

1.4.7 The candidate gene approach 51

1.5 SPECIFIC AIMS AND OBJECTIVES OF THIS THESIS 54

1.6 CHAPTER SUMMARY 56

This thesis describes the results of genetic mapping of agronomic traits in willows grown for biomass production through the application of molecular marker techniques. This chapter introduces willows in the context of their use for biomass production, outlines the main molecular marker techniques available and then reviews how molecular markers can contribute to crop improvement programmes. The chapter concludes with the specific aims and objectives of the thesis.

1.1 The genus *Salix*

Willows (genus *Salix)*, and also poplars and aspens (genus *Populus*), belong to the family *Salicaceae*. The genus *Salix* is very heterogeneous, comprising over 300 species (Stott, 1984) with considerable variation in size and growth form. Three main subgenera are recognised (Table 1.1), although below this level of classification considerable confusion exists in the taxonomy. However, one of the most widely accepted classifications is that of Skvortsov (1968).

Table 1.1. The three main subgenera of the genus *Salix*.

Subgenus	Characteristics	Examples
Salix (true willows)	Upright, pendulous or semi-pendulous trees and large shrubs with narrow, serrate leaves	*S. alba, S. babylonica, S. fragilis, S. nigra*
Caprisalix (osiers & sallows)	Shrubs and small trees with great variation in leaf form	*S. caprea, S. cinerea, S. aurita, S. viminalis*
Chamaetia (dwarf willows)	Small mountain or Arctic shrubs with rounded or blunt leaves	*S. reticulata, S. herbacea, S. retusa, S. myrsinites*

The genus *Salix* is thought to have originated in what are now the mountains of Eastern Asia. This is supported by apparent links in this region between the *Chosenia,* which have many basic *Salix* features, and some primitive members of the genus *Populus*. The genus is thought to have spread over considerable distances following the most recent ice age, when glaciers linking the continents of the northern hemisphere melted to form vast rivers that carried sediment across Asia, Europe and North America. Being one of the earliest colonisers of glacial sediment, *Salix* thrived during this period and became well distributed. As the glaciers receded, *Salix* populations remained widely disseminated, mainly in temperate and Arctic regions of the northern hemisphere. Today, willows can be found, through natural distribution or introduction, in most parts of the world.

In willow, the haploid chromosome number is 19, but many species are tetraploid and higher ploidy levels are commonly found (Håkansson, 1955; Suda and Argus, 1968, Büchler, 1985, 1986). Willows are dioecious, with individuals bearing either all male- or all female-flowered catkins. Hybridisation between *Salix* species occurs frequently in nature (Nilsson, 1954; Argus, 1974), with the exception of hybridisation between members of the *Salix* subgenus and those of the *Caprisalix* and *Chamaetia* subgenera, which does not generally occur (Stott, 1984). Many factors such as dioecism and the varying flowering times associated with different species, also affect hybridisation success (Newsholme, 1992). Most hybrids within the subgenera are fertile and can cross with other species/hybrids, further complicating the problems associated with classification and identification. Common hybrids in the UK include those between *S. alba* and *S. fragilis* (true willows) and those between *S. repens* and *S. aurita* (sallows). The relative ease with which hybridisation occurs in nature, and the existence of a large number of willow species, has resulted in tremendous diversity within the genus.

1.2 Short Rotation Coppice (SRC) willow as a biomass crop

Increased energy production from renewable resources is currently a high priority in Europe as a result of dwindling fossil fuel reserves and the need to produce energy in ways that are non-polluting. The generation of energy from biomass sources can contribute towards a solution. Ambitious goals have been set for increasing energy production from renewables within the European Union, including a commitment to increase the energy production from biomass from 3 % in 1997 to 8.5 %, or 90 million tonnes of oil equivalent (mtoe) per year, by the year 2010 (European Commission White Paper; Anon, 1997). While agricultural residues, such as those from forestry, straw and poultry litter, may provide a potentially valuable biomass resource, energy crops grown specifically for this purpose will be required to meet this target. It is estimated that planting biomass crops on 7.1 % of all EU agricultural and forestry land (10 million hectares) will be necessary. Willow grown as Short Rotation Coppice (SRC) has been highlighted as a crop particularly suited to this purpose (Corbett and Britt, 1997).

SRC willows make an attractive biomass crop for a many reasons. SRC is a fast-growing, high production, low input crop, with minimal labour, fertiliser and chemical requirements. Plantations are clonally propagated and can be harvested every three to

five years, after which cut stools regenerate to provide another harvest and may continue to do so up to six times, giving a productive lifespan of 20-30 years. This provides a regular income to the farmer with faster returns than those associated with conventional forestry (Dawson, 1992).

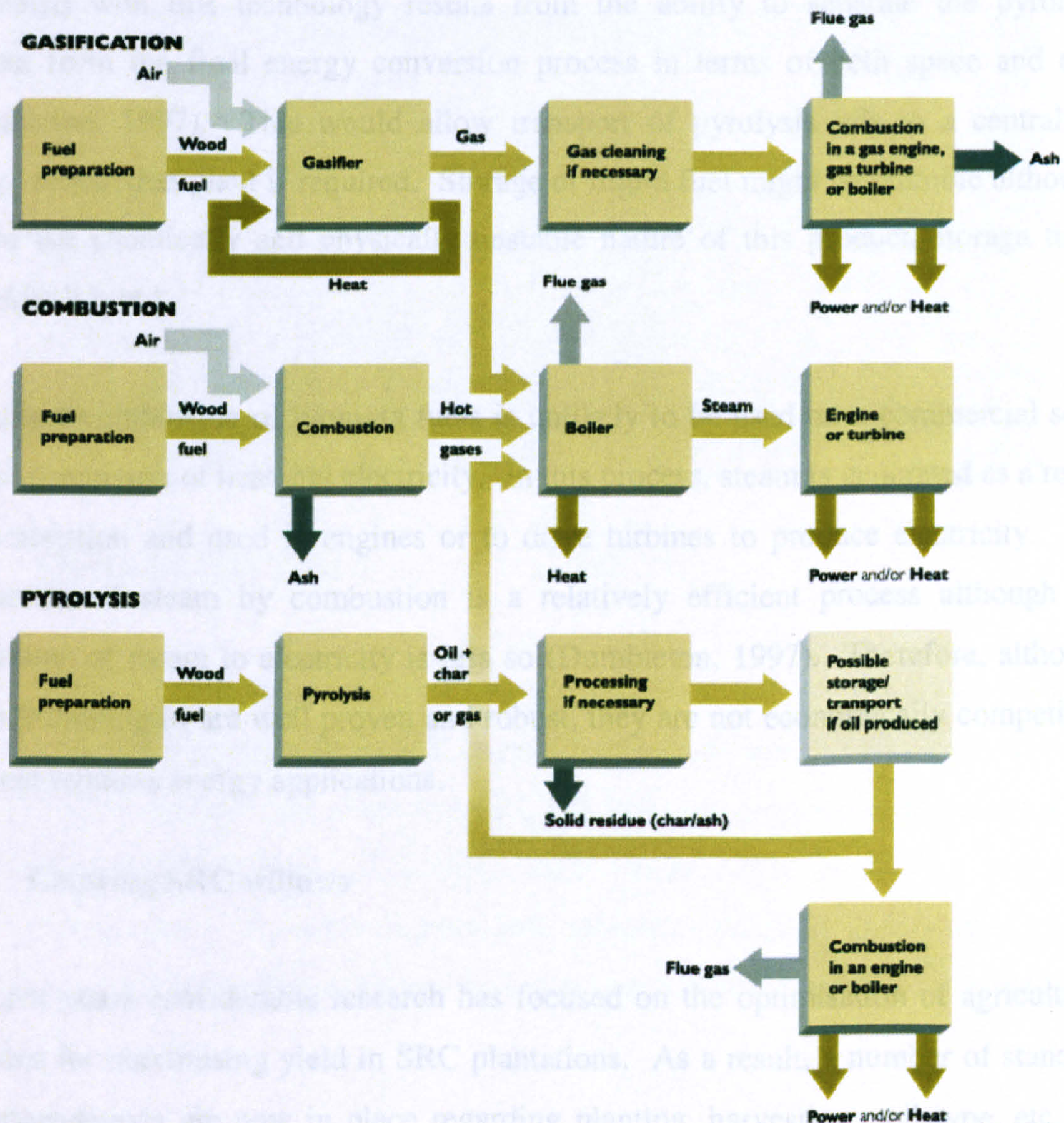
Furthermore, energy production from biomass willow is a non-polluting, closed-carbon system producing no net increase in the emission of greenhouse gases associated with global warming (Pitcher and Hilton, 1997). Carbon dioxide produced by the burning process is recycled via photosynthesis in stands growing to replace those harvested. As planting new trees increases carbon fixation in the soil, net emissions of carbon dioxide may even be negative (Hall *et al.*, 1997). The recurrent use of land may provide a more efficient and cost effective method for negating carbon emissions from other sources than the sequestration of carbon in new forests which may cease after 20 years (Hall *et al.*, 1997). This environmentally sound system has a high potential for public acceptance.

A further benefit associated with growing SRC willow as a biomass crop is the potentially positive impact on biodiversity, as SRC stands can provide shelter for a number of birds and mammals (Larsson, 1996). Also, in Britain, more insects live on willows than on any other tree species (Kennedy and Southwood, 1984). While many feed on willow, substantial damage needs to be occur before there is a negative impact on yield. The large numbers of insects in SRC plantations contributes to the biodiversity of intensively-farmed land (Tucker and Sage, 1999) and may also serve to attract greater numbers of insectivorous birds and small mammals.

1.2.1 Energy production from SRC

Three main technologies exist for the production of energy from SRC energy sources. These are gasification, pyrolysis and direct combustion (Figure 1.1).

Figure 1.1. The three main technologies used to generate heat and/or power from SRC.



Source: Short Rotation Coppice for Energy Production: Good Practice Guidelines, British Biogen. (http://www.britishbiogen.co.uk/gpg/srcgpg/srcgpgfront.htm).

At present, the favoured method is gasification, a thermochemical process involving the partial oxidation of the carbonaceous fuels at high temperatures and generally under high pressure (Nordin and Kjellström, 1996). A stable fuel gas results, consisting mainly of carbon monoxide and hydrogen with lesser amounts of methane and higher hydrocarbons such as ethane and ethylene (Dumbleton, 1997). This gas

can be cleaned if necessary and used in gas engines, boilers or in gas turbines to generate heat and/or power. Resulting flue gases are alkaline or neutral and are, therefore, environmentally friendly.

An alternative technology, pyrolysis, entails the thermal degradation of carbonaceous fuels in the absence of air or oxygen and results in the production of solid, liquid and gaseous products (Nordin and Kjellström, 1996). Resulting pyrolysis oil can be then be combusted in engines or boilers to produce heat and power. One advantage associated with this technology results from the ability to separate the pyrolysis process from the final energy conversion process in terms of both space and time (Dumbleton, 1997). This would allow transport of pyrolysis oils to a centralised energy production plant if required. Storage of liquid fuel might be possible although, due to the chemically and physically unstable nature of this product, storage times would be limited.

The direct combustion of biomass fuels is unlikely to be used on a commercial scale for the generation of heat and electricity. In this process, steam is generated as a result of combustion and used in engines or to drive turbines to produce electricity. The production of steam by combustion is a relatively efficient process although the conversion of steam to electricity is less so (Dumbleton, 1997). Therefore, although steam technologies are well proven and robust, they are not economically competitive for most biomass energy applications.

1.2.2 Growing SRC willows

In recent years considerable research has focused on the optimisation of agricultural practices for maximising yield in SRC plantations. As a result, a number of standard recommendations are now in place regarding planting, harvesting, soil type, etc. In general, SRC willows are grown in dense plots of between 10 000 and 20 000 stools per hectare (Figure 1.2), with spacings of 1.25 m x 0.75 m (Turnbull, 1997). Cuttings, normally between 15 - 20 cm in length, are planted by hand or specialised machinery during early spring months, typically in twin rows to facilitate subsequent mechanical harvesting. Growth is both rapid and vigorous, with one-year old stems reaching up to 5 metres in height. The standard rotation is three years although rotations of five years are not uncommon. Following each harvest, stems are generally chipped and dried

prior to use in energy production. The ability to store resulting woodchip for future use provides an advantage over other renewable energy sources, such as wind, where this is not possible.

Figure 1.2. A SRC willow plantation.

1.2.3 Threats to SRC yield

Pests, diseases and competition from weeds are important factors that potentially threaten to decrease yields from SRC crops (Tucker and Sage, 1999). Such pressures are of critical consideration, given that SRC is a low cash crop and economic viability is reliant upon obtaining consistently high yields while employing low-input agricultural practises. In response, numerous research studies are underway to further understand these potential problems and identify suitable methods for sustainable management.

1.2.3.1 Diseases of SRC willow

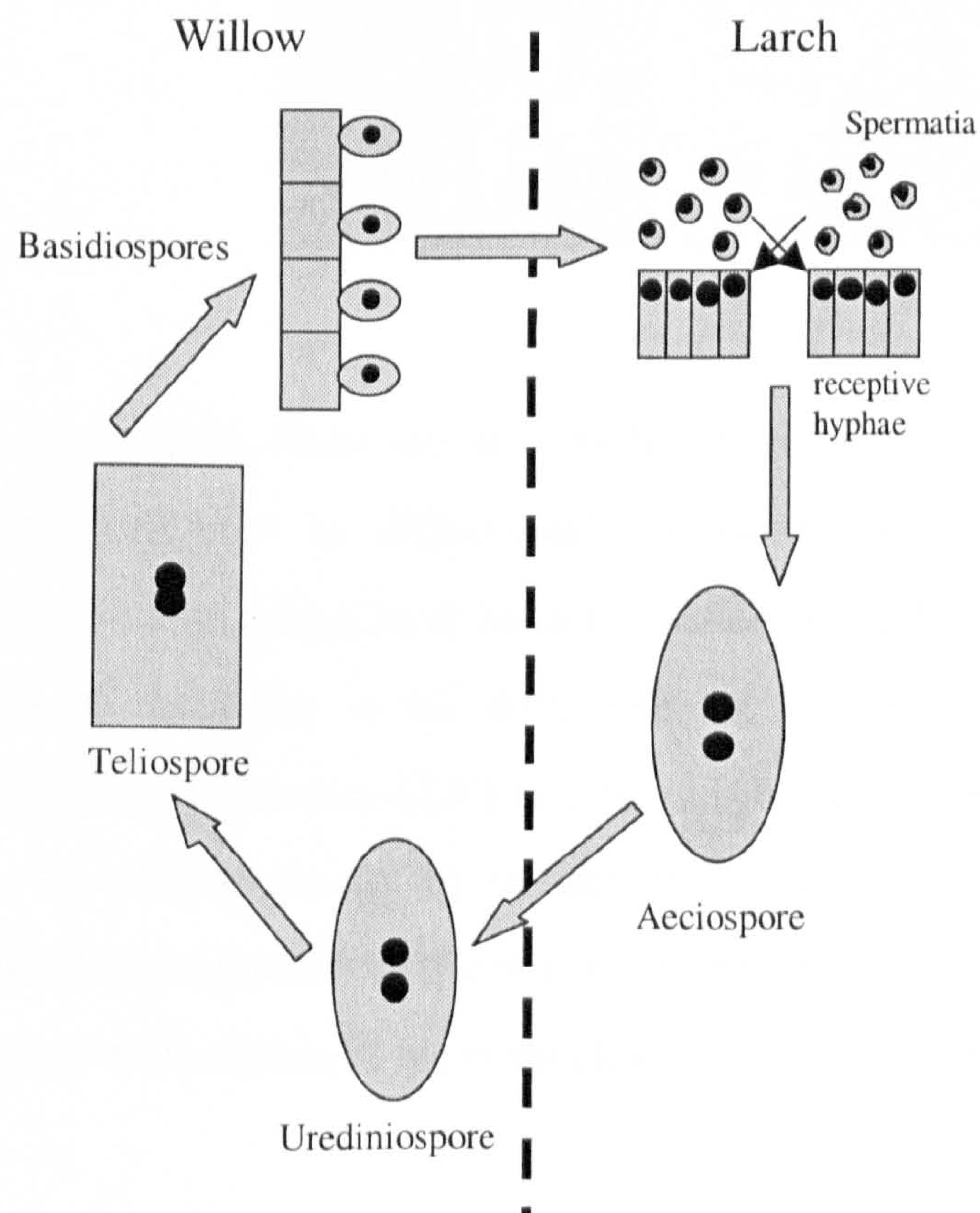
1.2.3.1.1 Rust diseases

Rust diseases are caused by fungi of the order *Uredinales,* a diverse group of plant pathogens comprising 5 000-6 000 species with highly specialised pathogenicities (Pei *et al.*, 1999). Rust fungi are found globally, with the exception of Antarctica. They infect almost all plant types and can be economically devastating to many crops.

Rusts caused by *Melampsora* spp. pose the greatest threat to SRC willow productivity in the UK and Europe (Pei *et al.*, 1997). In severe cases, premature defoliation can occur, at times leading to a predisposition to secondary infection and, ultimately, potential yield of up to 40 % (Pei *et al.*, 1997). In the most susceptible clones, such as *S. burjatica*, Korso, complete crop failure can occur (Dawson and McCracken, 1994).

Seven species of *Melampsora* have been described on willows in the UK (Pei *et al.*, 1993). *M. epitea* and *M. capraearum* are common on shrub willow species favoured in SRC plantations. Both have complex life cycles typical of the heteroecious *Melampsora* rust fungi (Figure 1.3), comprising five spore stages and involving an alternate host, mainly European larch (*Larix decidua).*

Figure 1.3. The life cycles of *Melampsora epitea* and *M. capraearum.*



Source: Pei *et al.*, 1997.

Urediniospores harboured in uredinia (pustules) are the cause of many cycles of disease during the growing season and can be observed on leaves during late spring through summer (Figure 1.4). During autumn, teliospores form on infected leaves and over-winter until germination and the production of basidiospores in spring. Basidiospores then infect the alternate host larch and produce spermatia and receptive

hyphae. Fertilisation is followed by the formation of aeciospores and the cycle begins once more.

Figure 1.4. A willow leaf infected with rust.

Within SRC plantations, rusts are also highly variable within species. For example, based on pathogenicities to differential willow hosts, 14 distinct pathotypes of *Melampsora epitea* var. *epitea* have been identified in the UK (Pei *et al.*, 1997). These pathotypes can be assigned to the form species: *larici-epitea typica* (LET), *larici-daphnoides* (LD), *larici-retusae* (LR) and *ribesii-purpurea* (RP) (Pei *et al.*, 1996). The form species have been shown to be genetically distinct by *in vitro* hybridisation experiments in which genetic barriers prevent the free exchange of genetic material between isolates of the different form species (Pei *et al.*, 1997).

The emergence of new pathotypes with novel virulences has been observed over the past decade (McCracken, 2001). For example, the cultivar *S. x mollissima*, "Q83", was rust resistant prior to 1992 when a new pathotype, LET4, emerged. By 1993, this pathotype was well established and was attributed to be the cause of severe infection on "Q83" at a number of UK sites. Similarly, *S. burjatica* "Germany", a previously resistant cultivar, succumbed to severe rust infection in 1994 as a result of the emergence of the new pathotype, LR3 (Pei *et al.*, 1997). Selection for new virulences

may have occurred as a result of the establishment of an increasing number of willow plantations, consisting of mainly rust resistant clones, in recent years.

1.2.3.1.2 Disease management

In high value crops, such as wheat, the use of fungicides to control disease is economically justified. Furthermore, the seasonal changing of crop lines contributes to effective disease control. In SRC willow crops, the use of fungicides is not economically viable, and stands must remain productive for up to 25-30 years following establishment, with little or no intervention. Consequently, the threat from disease has serious implications regarding the future success of SRC as a crop, and alternative management strategies are urgently required. Studies have shown that the use of clonal mixtures in SRC plantations may be effective in reducing the impact of rust disease (McCracken and Dawson, 1997; Hunter and Peacock, 2001; McCracken, 2001). This strategy has also been successfully employed for disease control in cereals although concerns have been expressed over the emergence of pathogen 'super-races' that are capable of infecting all clones. However, to date there is no evidence of this occurring in willow (McCracken and Dawson, 1997).

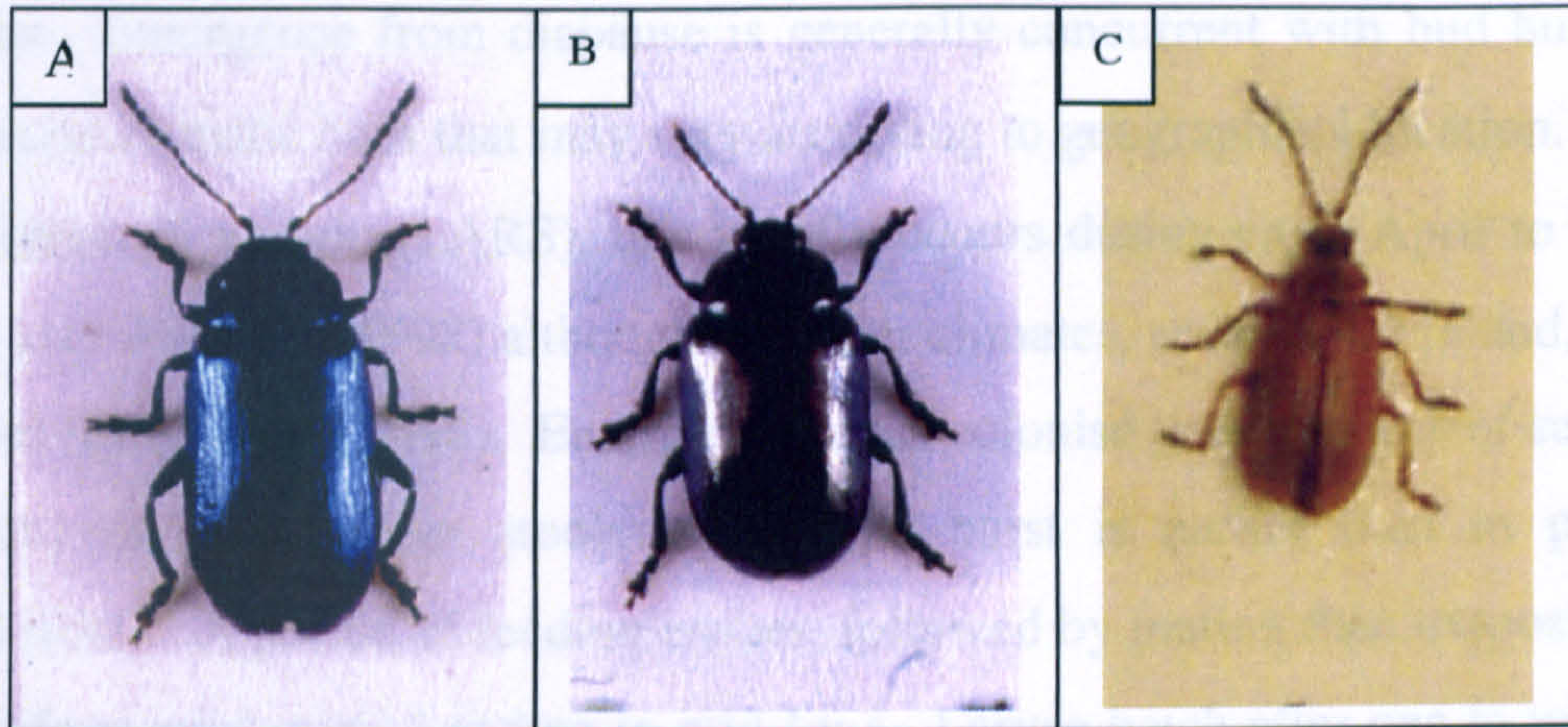
1.2.3.2 Invertebrate pests

SRC willows are known to provide a habitat for a large number of invertebrate species (Sage and Tucker, 1997) and their cultivation may prove beneficial in increasing biodiversity on farmland (Tucker and Sage, 1999). However, there is risk of some species becoming sufficiently numerous to be considered as pests.

1.2.3.2.1 Leaf-feeding willow beetles

In UK willow plantations, the most abundant invertebrates are herbivorous leaf beetles (Coleoptera: Chrysomelidae) (Sage and Tucker, 1997), of which, *Phratora vulgatissima* (the blue willow beetle), *P. vitellinae* (the brassy willow beetle) and *Galerucella lineola* (the brown willow beetle) have been consistently highlighted as the most damaging (Kendall *et al.*, 1996; Sage and Tucker, 1997; Wiltshire *et al.*, 1997) (Figure 1.5).

Figure 1.5. The three main willow beetle species that cause damage to SRC plantations – *Phratora vulgatissima* (A), *Phratora vitellinae* (B) and *Galerucella lineola* (C).



These beetles have also been reported as serious pests throughout Europe, in countries including Sweden (Wiren and Larsson, 1984), Finland (Kolehmainen *et al.*, 1995), France (Rowell-Rahier, 1984), Belgium (Soetens and Pasteels, 1994) and Ireland (Kelly and Curry, 1991). The level of defoliation caused by the chrysomelid beetles can be considerable and cause substantial yield loss (Sage and Tucker, 1997). Complete defoliation of a willow plantation by *P. vulgatissima* has been reported (Kendall *et al.*, 1996). Leaf damage caused by *P. vulgatissima* is shown in Figure 1.6.

Figure 1.6. Leaf damage caused by *Phratora vulgatissima* – the blue willow beetle.

The life cycles of *P. vulgatissima*, *P. vitellinae* and *G. lineola* have been studied across a number of sites and are reported to be very similar (Hutchinson and Kearns, 1930a, b; Kelly and Curry, 1991; Kendall and Wiltshire, 1998; Rank *et al.*, 1998). In brief, adult beetles aggregate and hibernate in winter in sheltered sites local to SRC

plantations. Such overwintering sites include loose bark and cracks on trunks and branches of hedgerow trees such as hawthorn, maple and willow, under moss, ground debris, leaf litter and in weed vegetation on the plantation floor. Stem cankers, caused by stem-infecting rust, have also been shown to shelter hibernating *P. vulgatissima* aggregates. Emergence from diapause is generally concurrent with bud burst and is dependent on climatic cues that may vary according to geographical location. At Long Ashton Research Station (LARS), this usually occurs during early April to mid May (Kendall and Wiltshire, 1998) although in colder climates, such as in Finland, this may be delayed (Rank *et al.*, 1998). Emerging beetles colonise young leaves of salicaceous trees, particularly on uncut stools where bud burst is earlier than in previously coppiced stools. A period of feeding ensues, followed by mating then oviposition over a three to four week period ending in mid June. Larvae hatch after one to two weeks, develop through three larval instars and pupate before a new adult emerges after a total development time of approximately 65 days. Ecolded adults then feed on the crop prior to hibernation in autumn. Thus, throughout the growing season there is always a life stage present in the crop capable of causing sustained defoliation. While *P. vulgatissima* are univoltine in the UK, with only one generation per year, *P. vitellinae* and *G. lineola* are capable of producing a complete or partial second brood in late summer if climatic conditions are suitable. The frequency of occurrence of a second generation may affect defoliation levels incurred throughout the season and may, therefore, have significant economic implications. In Sweden and Finland, where the climate is colder, *P. vitellinae* are always univoltine (Denno *et al.*, 1990).

Studies performed throughout Europe indicate that there is considerable variation in the susceptibility of willow and poplar cultivars and species to feeding damage by willow beetles (Rowell-Rahier, 1984; Wiren and Larsson, 1984; Tahvanainen *et al.*, 1985; Kelly and Curry, 1991; Soetens and Pasteels, 1994; Kendall *et al.*, 1996; Wiltshire *et al.*, 1997). These beetles have been shown to discriminate between host species and consistently feed on some cultivars in preference to others. In addition, there is evidence that the preferred host choice differs between the willow beetle species, i.e. *P. vitellinae* prefers to feed on different clones to *P. vulgatissima* and *G. lineola*, which have been shown to have similar feeding preferences in some studies (Wiltshire *et al.*, 1997). Various chemical and morphological characteristics of willow leaves have been implicated in host-plant selection including water and nutrient status of the leaf (Mattson, 1980) and leaf texture (Rowell-Rahier and Pasteels, 1982).

However, the chemical composition of secondary metabolites in the leaves has been implicated as the most important factor influencing feeding preference, with phenol glucosides such as salicortin, salicin, saldroside and tremulin playing key roles (Tahvanainen *et al.*, 1985; Kolehmainen *et al.*, 1995). Species-specific differences in concentration and composition of these compounds have been reported for different willows (Thieme, 1971; Palo, 1984), leading to the widely accepted hypothesis that these metabolites influence the host-plant selection of the beetle pests. For example, in laboratory and field-based studies, *P. vulgatissima* and *G. lineola* have been shown to prefer feeding on willows containing low concentrations of phenol glucosides, such as *S. viminalis* and several hybrids of *S. viminalis, S. aurita, S. caprea* and *S. cinerea.* Willows containing moderate to high concentrations of these secondary metabolites, such as *S. burjatica, S. dasyclados, S. purpurea* and *S. triandra,* were least preferred (Wiltshire *et al.*, 1997). In contrast, the feeding preferences of *P. vitellinae* positively correlated with moderate to high concentrations of particular phenol glucosides, particularly salicin or salicortin, in host leaves (Pasteels and Rowell-Rahier, 1992).

1.2.3.2.2 Willow beetle management strategies

The most obvious strategy for willow beetle management would be the use of insecticides. However, such an approach has limited economic value for this low-cash crop and application is difficult due to the dense nature of agroforestry plantations (Kendall *et al.*, 1996). Furthermore, the broad-spectrum activity of such chemical agents may have major environmental side effects (McCracken and Dawson, 1997) and negate the benefits associated with SRC as a habitat for wildlife and a potential source of beneficial organisms in agroecosystems.

Alternative approaches, aimed at decreasing the levels of damage inflicted by these pests, have been suggested. The first targets the hibernation stage of the willow beetle life cycle. It is believed that the removal of overwintering sites local to SRC stands may be beneficial in decreasing the number of beetles present for spring invasion. This may be aided by improved attention to crop hygiene, such as the removal of plant debris from the site and effective weed control. Furthermore, the establishment of willow plantations on sites that are well isolated from surrounding hedgerows and woodland may be beneficial.

A second proposed management strategy exploits differences in feeding preferences associated with different willow beetle species. It has been suggested that the selection of willow clones for a particular site should be influenced by the feeding preferences of the dominant willow beetle species in that location (Sage and Tucker, 1997). Furthermore, as with reduction of rust disease, the use of clonal mixtures in SRC plantations has been highlighted as a promising method for reducing damage sustained through beetle feeding (Peacock *et al.*, 2001).

Although considered to some degree within the willow breeding programmes (Lindegaard and Barker, 1997), resistance to willow beetle herbivory has not been a primary focus for selection, mainly due to the limited damage caused by these pests at present. However, as more SRC plantations are established at a greater number of sites, problems relating to insect pest damage have the potential to increase. Given the contrasting feeding preferences of the different willow beetle species, it may be possible to breed improved varieties that incorporate the naturally occurring resistances derived from different willow species into single highly resistant varieties.

1.2.3.2.3 Other insect pests

In addition to the chrysomelid beetles, several other insect species are capable of causing damage to SRC plantations. In Sweden, galling midge species have been reported to be the most serious pests of SRC plantations (Sage and Tucker, 1997). These insects also occur in the UK but have not yet been implicated as the cause of severe coppice damage. However, in a Swedish study, the leaf roll gall midge (*Dasineura marginemtorquens*) was shown to cause 38 % growth loss in willow (Glynn, 1996). The aphids *Tuberolachnus salignus* and *Pterocomma salicis* are also potential pests commonly observed in SRC plantations (Collins, 1999) although, to date, the extent of damage caused by these organisms remains unclear.

1.2.3.3 Mammalian herbivores

Grazing by mammalian herbivores such as deer, hares and rabbits can also cause severe SRC yield losses. This is of particular consequence in small, isolated, newly established coppices at sites inhabited by significant populations of these animals. In larger plantations, the affect of grazing is of lesser importance (Larsson, 1996).

Differences in palatability have been described between willow cultivars (Larsson, 1996), indicating that 'resistance' to mammalian herbivory may potentially be incorporated into willow breeding regimes.

1.2.4 Breeding of improved biomass willows

As the commercial success of SRC willow plantations is highly dependent on yield it is crucial that highly productive, pest and disease resistant varieties are developed. Furthermore, in order to meet renewable energy targets as outlined in Section 1.2, new and improved high yielding willow varieties will be required.

1.2.4.1 Historical overview

Willows have been cultivated for centuries for use in basketry and to make furniture (Stott, 1992). However, many varieties used for such purposes are not ideally suited for use in energy plantations where the importance of total above-ground biomass is paramount. To address this problem, a breeding programme aimed specifically at the production of varieties suitable for biomass uses was initiated in 1987, by the Swedish plant breeding company, Svalöf Weibull AB (Åhman and Larsson, 1994). Later, in 1996, the European Willow Breeding Partnership was formed and sited at LARS, bringing together both the Swedish and UK efforts.

Initially, these breeding programmes centred around a number of candidate species identified as having good potential for biomass use (Stott, 1984), the majority of which were basket willows of the *Caprisalix* sub-genus. However, in order to increase genetic diversity in the breeding populations, additional material from wider geographical locations has been incorporated more recently, e.g. *S. schwerinii* and *S. dasyclados* clones from Siberia (Larsson, 2001). In addition, collections such as the UK National Willow Collection (NWC) held at LARS, which contains 1300 clones comprising 120 different species, have been used as a valuable source of additional genetic variation. A second germplasm collection is maintained by Svalöf Weibull AB in Sweden.

The genetic improvement of SRC willow as a crop has been both significant and rapid in recent years. This is not surprising given the broad genetic base of available starting material and that the breeding of willows for biomass is a relatively novel endeavour. The first new varieties aimed specifically at biomass use were bred in Sweden by Svalöf Weibull AB and were released during the early nineties. These varieties, namely Orm, Rapp, Ulv, Jorr and Jorrun, were reported to produce yields 15 – 20 % higher than those achieved with earlier non-bred clones (Larsson, 1996). Since then, the release of new and improved varieties has continued, with clones such as Björn and Tora (bred in Sweden) and Ashton Stott (bred in the UK) being the most promising to date. Yield increases of approximately 50 % have been reported more recently (Larsson, 2001).

1.2.4.2 Breeding objectives and strategy

In general, the Swedish and British breeding programmes have common goals, as outlined in Table 1.2. However, some differences exist due to climatic variation and dissimilar pest and disease pressures at specific geographical locations. For example, varieties produced for colder climates, such as in northern Sweden and Finland, must show a greater tolerance to frost than those produced for UK conditions.

Table 1.2. Fundamental objectives of British and Swedish biomass willow breeding programmes.

Yield	1	Increase dry matter yield/hectare
	2	Select tallest clones with either few, thick or many, thin shoots per stool
Disease and pest resistance	3	Improve resistance to willow rust
	4	Improve resistance to willow beetles
	5	Improve resistance to *Dasineura* spp.
	6	Select clones less palatable to mammalian herbivores
Harvesting ability	7	Improve growth form. Select straight rods with few side branches
Frost tolerance (Sweden)	8	Improve tolerance to frost for northern climates

Adapted from: Lindegaard and Barker, 1997.

Several crossing strategies have been employed within breeding programmes to achieve the desired objectives. These include intraspecific crosses, such as those between *S. viminalis* clones that produced the early Swedish varieties Orm, Rapp, Ulv, Jorr and Jorrun. However, gains associated with this type of cross may be limited due to the narrow genepool involved. Alternatively, interspecific crosses may be

performed and may result in heterosis (hybrid vigour), where all desirable traits of the parents are incorporated into the progeny resulting in offspring that exceed the expected performance of the parental mean. An example of this phenomenon resulted from a cross between *S. viminalis*, Bowles Hybrid, and *S. burjatica*, Korso, that produced Ashton Stott, a clone that consistently performs better than the mean performance of the two parents (Lindegaard and Barker, 1997). Crosses between superior native clones and exotic relatives have also proved successful, illustrated by the creation of the high-yielding and disease-resistant varieties, Björn and Tora. More speculative interspecific crosses involving species previously unused for SRC, and those that produce triple hybrids, have also been attempted.

1.2.4.3 Breeding practicalities

Several characteristics of willows make them particularly amenable to breeding. Firstly, willows generally reach sexual maturity in the second year of growth. This is relatively early in comparison to other tree species (Lindegaard and Barker, 1997). In addition, the dioecious nature of willow means that pollen can be transferred directly to the female plant without any need to guard against self-pollination. This is in contrast to monoecious crops, such as maize, where male components of a flower must be emasculated to prevent this occurring. Furthermore, as willows can be propagated from cuttings, crosses can be made using cut material in the glasshouse.

At LARS, crossing is usually performed in January or February, at which time shoots bearing flower buds are transferred to the glasshouse and grown in glass jars containing water. At this stage male and female cuttings are isolated from one another. Pollen is collected from the males and applied to receptive female stigmas using a small brush. Seed-set is normally observed after a period of three to six weeks. When mature, seeds are threshed and subsequently sown in trays containing peat and sand. When a height of approximately 3 cm is reached, seedlings are pricked out and transferred to peat-filled polystyrene cell modules. At a height of 10 cm, seedlings are then transferred to trays containing soil and slow-release fertilizer and re-located outside to the nursery.

1.2.4.4 Current breeding and selection schemes

Within the LARS breeding programme, initial selections are performed in the nursery on the basis of easily observed traits such as resistance to *Melampsora* rust, height and growth form. Selected individuals are then planted in two field-based observation trials that differ in planting design. Further assessments are then performed focusing on pest and disease resistances and rough estimates of yield obtained from the measurement of potential yield components such as stem height, stem diameter and number of shoots per stool. Genotypes that perform well according to the selection criteria are then planted in yield trials, which also contain reference clones for ease of comparison and early identification of elite individuals that may eventually be named as new varieties. A more comprehensive description of the LARS breeding scheme is provided by Lindegaard and Barker (1997) and Lindegaard (2002).

1.2.4.5 Future breeding prospects

Given the relative infancy of biomass willow breeding programmes in comparison to those for more conventional food crops, it is likely that the potential for producing superior genotypes is immense. Moreover, the vast levels of genetic diversity within *Salix*, and the ease with which hybridisations can be achieved, suggest that there may still be much to gain from the incorporation of previously untapped germplasm sources. However, at present, breeding efforts are hampered by a paucity of information regarding the genetic basis of agronomically important traits and inefficient and time-consuming selection procedures (which are often based on simultaneous selection for numerous traits). Despite these potential limitations, significant progress has resulted from the both Swedish and UK-based programmes in recent years (Larsson, 2001; Lindegaard *et al.*, 2001). However, the efficiency of future breeding efforts is amenable to significant improvement by the effective coupling of conventional breeding techniques with molecular marker technologies.

1.3 Genetic markers

Genetic markers are heritable chromosomal landmarks that can be monitored and used in genome analysis and inheritance studies. There are two main categories: morphological, and biochemical or molecular markers.

1.3.1 Morphological markers

Morphological markers are based on the visualisation of simply-inherited phenotypic traits under the control of a single locus. Such markers have been extensively used in classical studies of heredity. It was through the use of carefully selected morphological markers that Mendel conducted his experiments with garden peas and deduced the fundamental laws of genetics. Morphological markers also enabled early geneticists to recognise and investigate linkage, sex-linkage, gene interactions and the difference between discontinuous and continuous (or quantitative) variation and to study genes in populations. However, the number of potential morphological markers available to geneticists is limited and, to be useful, markers of this type must be expressed and distinctly recognisable over a range of environments. Also, their expression may be further complicated by genetic factors, such as epistatic or pleiotropic interactions (Staub *et al.*, 1996). Furthermore, morphological markers often have such large effects on phenotype that their use is limited for many applications, e.g. in plant breeding programmes (Tanksley *et al.*, 1989).

1.3.2 Molecular markers

The use of molecular markers revolutionised genetics principally by providing a new supply of character differences that could be detected in laboratory assays, rather than through analysis of growth or morphological characteristics. There are two distinct classes: those based on detection of difference in proteins and those based on detection of variation in DNA.

Initial molecular marker studies were protein-based, focussing on isozymes. These are variant forms of the same enzyme that differ in charge, and can, therefore, be separated by electrophoresis on starch gels (Markert and Moller, 1959). The advent of this technology greatly enhanced genetic studies at the time, although these techniques

have a number of associated disadvantages. Firstly, biochemical markers are products of gene expression and may, therefore, be subject to post-translational modifications that can often restrict their usability (Staub *et al.*, 1982). Furthermore, sensitivity to environmental factors and a general lack of isozyme loci may also restrict their use (Staub *et al.*, 1996).

The use of DNA-based molecular marker systems circumvents many of these problems. Such markers are not products of translation and are subsequently unaffected by environment. Also, the number of potentially useful markers is immense and is not generally a limiting factor for most applications (Kumar, 1999). Furthermore, DNA is present in virtually all cells of organisms, it can be recovered from both living and dead tissue, and, in many cases, only nanogram amounts are required for analysis.

The immense growth in application of molecular marker over the past decade has yielded many significant developments in this field, with the advent of many novel marker technologies and significant advances in their application to a variety of genetic disciplines (Bachmann, 1994; Lee, 1995, Karp *et al.*, 1996; Karp *et al.*, 1997; Lee, 1998; Kumar, 1999). Numerous DNA marker systems are now available to the scientific community, the most common of which are described below. Particular attention is afforded to microsatellite and amplified fragment length polymorphism (AFLP) markers, which have been most intensively employed in the research described later in this thesis.

1.3.2.1 Restriction Fragment Length Polymorphism (RFLP)

The first widely used DNA-based molecular marker systems relied on hybridisation methodologies to detect polymorphisms. One such marker class is Restriction Fragment Length Polymorphism (RFLP), a technology which takes advantage of naturally occurring variation in restriction sites (Sambrook *et al.*, 1989). This variation occurs as a product of stable, heritable, mutations that may cause loss or gain of restriction sites, providing a rich source of polymorphisms that can be exploited in genetic studies. As the genomes of most plants contain between 10^8 and 10^{10} nucleotides, mutations in a small proportion of sites can give rise to a large number of putative DNA markers (Paterson *et al.*, 1991). To detect polymorphisms, genomic

DNA is first digested with restriction enzymes to produce a pool of differently sized DNA fragments. The composition of the resulting pool is dependent on the number and genomic distribution of restriction sites in the original sample. The fragments are then separated by gel electrophoresis, blotted onto a nitrocellulose or nylon filter, and probed with small lengths (usually 500 - 3000 bp) of radio-labelled cloned genomic DNA or cDNA. The location of the probes on the filters is then visualised via autoradiography revealing a banding pattern that can be used in a variety of genetic studies.

Markers based on RFLP technology are co-dominant, enabling homozygotes to be distinguished from heterozygotes. A further advantage is the high level of reproducibility between laboratories. Disadvantages associated with this marker type stem from the requirement for sufficient numbers of probes and the time-consuming blotting and hybridisation steps that cannot be easily automated (Mazur and Tingey, 1995). Also, relatively large amounts of high quality DNA are required.

1.3.2.2 Arbitrary priming marker classes

The development of the Polymerase Chain Reaction (PCR) (Saiki *et al.*, 1988) brought about the rapid development of a range of new marker technologies that were able to surmount some of the technical limitations of RFLP. Several of these markers systems employ arbitrary priming in PCR to detect polymorphisms, i.e. one or more synthetic oligonucleotides of arbitrary sequence are used to amplify specific, but unknown, sites in a genome. Examples of markers systems based on arbitrary priming include Randomly Amplified DNA Polymorphism (RAPD) (Williams *et al.*, 1990), Arbitrary Primed-PCR (AP-PCR) (Welsh and McClelland, 1990) and DNA Amplification Fingerprinting (DAF) (Caetano-Anollés *et al.*, 1999), which differ from one another essentially in the length of primer, primer/template ratio, the gel matrix used and the method of visualisation. Such strategies have the advantage that no prior sequence information is required for the design of specific primers and they can, therefore, easily be applied to any organism. However, the resulting markers are dominant and heterozygotes cannot be identified. Furthermore, problems relating to reproducibility between different laboratories have been reported (Jones *et al.*, 1997). Also, problems relating to scoring may arise as amplification products of equal electrophoretic mobility will co-migrate to equivalent positions on a gel.

1.3.2.3 Amplified Fragment Length Polymorphism (AFLP)

A number of the disadvantages associated with RFLP and arbitrary priming methodologies were overcome with the development of Amplified Fragment Length Polymorphism (AFLP) (Zabeau and Vos, 1993; Vos *et al.*, 1995) essentially a hybrid of the two earlier strategies. As with RFLP, genomic DNA is first subjected to restriction digestion. In AFLP, simultaneous digestion by two restriction enzymes (usually a rare and a frequent cutting enzyme) is the norm. Adapters are then ligated to the restriction fragment ends to facilitate amplification with adapter-homologous, radio-labelled or fluorescently-tagged primers. Since the number of amplified fragments is potentially immense, the addition of two or three arbitrary selective nucleotides to the 3' end of the primers serves to reduce complexity by allowing amplification of only a subset of fragments. Resulting PCR products are separated by denaturing polyacrylamide gel electrophoresis and visualised using autoradiography or fluorescence-based detection systems, as appropriate. Size differences of one base pair can be resolved by these techniques.

AFLPs are highly reproducible between different laboratories (Jones *et al.*, 1997) and their capacity to generate large numbers of markers, with minimal effort and in relatively short timeframes, has made them the marker of choice for many genetic studies. Furthermore, as with arbitrary priming methods, no sequence information is required for their production. Lesser starting amounts of DNA are required in comparison to RFLPs, although slightly more is required compared with RAPDs. However, AFLP markers, like RAPDs, are dominant in nature and the information they provide may be limited for certain applications. AFLPs are also subject to problems associated with co-migration of super-imposed bands of equivalent electrophoretic mobility, although to a lesser degree than RAPDs which utilise agarose rather than denaturing polyacrylamide gels.

1.3.2.4 Sequence-based PCR markers

In contrast with arbitrary priming techniques, sequence-based PCR markers require prior knowledge of the sequence at the target site such that amplification can be target –specific. Comparison of the amplification product(s) in different organisms allows DNA polymorphisms in the target sequence to be detected. Sequence data of this kind is the most informative marker data that can be achieved, however, to provide a

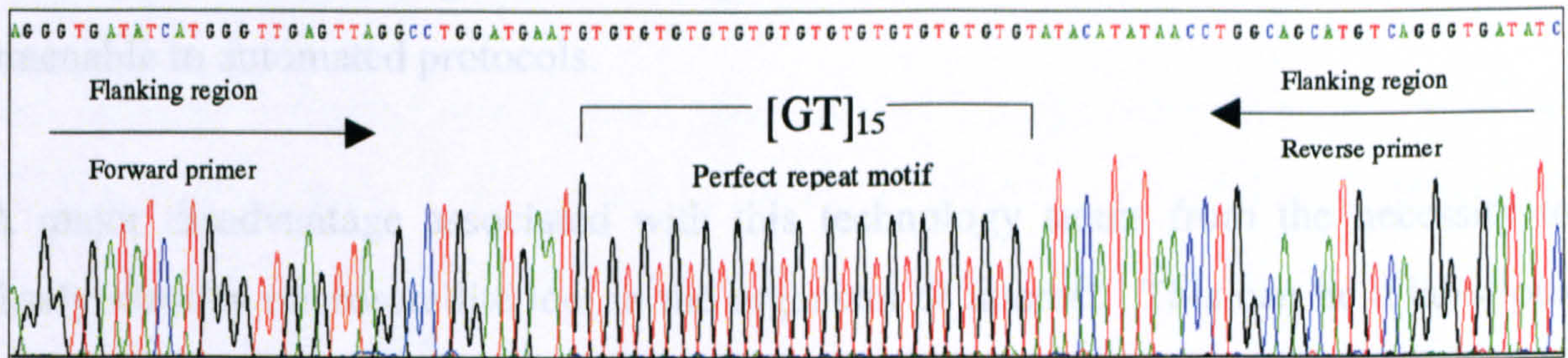
sufficient number of markers for use in genetic studies, it is necessary to have sequence information for a large number of genomic regions and, until recently, such information was not available. Advances in DNA sequencing and the undertaking of entire genome sequencing projects in several organisms, including humans, the fruit fly *Drosophila*, the weed plant *Arabidopsis*, rice, a nematode and several bacteria, have completely overturned this situation and there is now a constant flow of genome sequence information appearing in databases. Moreover, genomic initiatives throughout the world are yielding data on expressed regions of the genomes of a range of organisms and are assigning functions to an increasing number of sequences. Because markers of this kind are detected by amplification of target sequences they can be transferred from genome to genome, provided that mutations do not prevent primer annealing. This is significant, since the information from one genome (which may be a model organism) can be exported for use in another (which may be a specific crop). Unfortunately, at the onset of this project no genome sequence information was available for a tree crop. Thus, sequence-based markers in the current work have mostly been restricted to the use of hypervariable regions (microsatellites). However, efforts are now underway to sequence the poplar genome (Anon, 2002) which has direct relevance to the present work, given the close taxonomic relationship between willows and poplar and a number of genomics initiatives have also recently been initiated in poplar. It is anticipated, therefore, that future work in willow will utilise to a much greater extent sequence-based markers in genic regions.

1.3.2.4.1 Microsatellite markers

Microsatellites (Litt and Luty, 1989), also known as Simple Sequence Repeats (SSRs) (Jacob *et al.*, 1991), Simple Sequence Length Polymorphisms (SSLP) (Tautz, 1989) or Short Tandem Repeats (STR) (Edwards *et al.*, 1991), are DNA motifs consisting of tandemly repeated mono-, di-, tri-, tetra- or penta-nucleotide units (Powell *et al.*, 1996).

Repeat units longer than six base pairs (bp) are termed minisatellites (Jeffreys *et al.*, 1995). In the genomes of both plants and animals di-nucleotide repeats are most prevalent (Gupta *et al.*, 1996) (Figure 1.7).

Figure 1.7. Sequence electropherogram of a microsatellite locus showing a perfect di-nucleotide $[GT]_{15}$ repeat motif flanked by unique sequence to which primers can be designed for interrogation of size polymorphisms by PCR.



Repeat motifs may be 'perfect' or 'pure' having no interruptions to the repeat motif, 'imperfect' with non-repeat units present within the microsatellite or 'compound' consisting of two or more different adjacent repeat unit motifs (Weber, 1990). Microsatellites are both abundant and uniformly distributed throughout eukaryotic genomes, although there may be under-representation in coding regions. In contrast, minisatellites are generally confined to telomeric regions (Gupta *et al.*, 1996).

Microsatellite polymorphisms manifest as differences in the number of repeat units present at a given locus. Mutations that give rise to such variation are thought to be a consequence of slipped–strand mispairing during DNA replication (Schlötterer and Tautz, 1992) and occur at a greater frequency than those observed in non-repetitive sequences (Kashi *et al.*, 1997). The high mutation rate is consistent with the high frequency with which new alleles appear (Jeffreys *et al.*, 1988) and in the highly polymorphic nature of microsatellites between species and populations (Tautz, 1989).

Detection of microsatellite polymorphisms is generally achieved by direct amplification of the target locus from total genomic DNA followed by resolution of resulting PCR products by agarose or, more commonly, denaturing polyacrylamide gel electrophoresis. This approach exploits the phenomenon that, whilst a given microsatellite sequence may be present at several sites throughout a genome, those sequences flanking the microsatellite are usually unique and facilitate the design of locus-specific primers (Figure 1.7) for interrogation of a single, target microsatellite locus.

Resulting markers are co-dominant and highly informative, providing information on heterozygosity that, in general, cannot be obtained with dominant marker systems such

as AFLPs and RAPDs. Further benefits associated with microsatellite markers arise from their high reproducibility and the ease with which they can be distributed between different laboratories as locus-specific primer sequences. In addition, only small amounts of starting DNA are required and microsatellite genotyping is highly amenable to automated protocols.

A major disadvantage associated with this technology stems from the necessity to firstly identify microsatellite loci in the organism of interest. This can be a laborious and expensive undertaking, generally involving the production small-insert, microsatellite-enriched genomic libraries. The development and optimisation of such protocols has been the focus of much attention in recent years, and has resulted in a variety of methods becoming available (e.g., Brenig and Brem, 1991; Ostrander, 1992; Karagyozov *et al.*, 1993; Nishikawa *et al.*, 1995; Edwards *et al.*, 1996). The transferability of microsatellite primer sequences between laboratories may help to circumvent this problem in some cases, especially in well-studied organisms such as farm animals, crop species and model organisms, where the publication of microsatellite primer sequences, via public databases, may be exploited. Furthermore, several studies have shown that microsatellite markers developed for a particular organism may be used to target homologous loci in related species or genera (Coltman *et al.*, 1996; Jarne and Lagoda, 1996; Gemmell *et al.*, 1997; Moncrief *et al.*, 1997; Hanley *et al.*, 2002). Such markers are also useful for direct comparative studies between species and genera. However, it should be noted that the proportion of microsatellite loci that may be cross-amplified in related species diminishes as evolutionary distance between the species or genus increases.

A further disadvantage of microsatellites arises from their single-locus nature which can present a potential bottleneck in large-scale genotyping programmes. This is mainly due to limitations associated with conventional screening methodologies, in which radio-labelled primers are used in PCR targeting a single locus followed by detection via autoradiography. While these protocols allow for the simultaneous amplification of multiple loci in a single PCR reaction, detection systems based on radio-isotopes do not permit the assignment of resulting alleles to a particular locus unless their sizes are known *a priori*. Such an approach is, therefore, unsuited to many genetic studies, e.g. genetic diversity studies involving natural populations. However, the development of fluorescence-based detection systems that allow simultaneous

interrogation of multiple loci (multiplexing) has significantly increased the throughput potential of microsatellite screening programmes.

1.3.2.4.2 Markers based on expressed sequences

The increase in random and targeted gene discovery programmes in recent years has led to a sharp increase in the amount of sequence information available. In particular, sequence data for vast numbers of Expressed Sequence Tags (ESTs) are now held in both public and private databases. These resources provide a rich source of information that can be exploited in the development of molecular markers based on expressed genes. Markers of this type have several advantages over those based on anonymous sequences and are now becoming standard, not only in the study of human genetic diseases, but also in plant and animal improvement programmes (Saier, 1998; Picoult-Newberg *et al.*, 1999). The first advantage associated with markers based on expressed genes is that if an EST marker is found to be genetically linked to a trait of interest, then it is possible that the mapped gene directly affects that trait. Second, ESTs that share homology with functional candidate genes (Section 1.4.7) may be considered to be good initial targets for mapping studies. Likewise, ESTs that are highlighted in analyses of differential gene expression, e.g. in microarray experiments (Richmond and Somerville, 2000), may also become targets for further study. Finally, as ESTs are generated from coding DNA, which is generally more highly conserved than non-coding DNA (Cato *et al.*, 2001), resulting markers are more likely to be transferable across more genetically diverse pedigrees and species than are markers based on non-coding regions, e.g. microsatellites, RAPDs and AFLPs etc. Therefore, these markers may be more suited to comparative mapping studies in which genome maps and QTL are aligned (Section 1.4.4). Furthermore, this phenomenon affords the possibility of exploitation of EST resources originally developed in model organisms in related, less-studied animals and plants. However, taking an alternative viewpoint this last consideration highlights one of the potential drawbacks associated with EST mapping, in that the associated high levels of conservation within genes may provide a bottleneck in the discovery of suitable polymorphisms for mapping. However, several approaches have been suggested to circumvent this problem, such as PCR-RFLP (Tragoonrung *et al.*, 1992), heteroduplex analysis (Fischer and Lerman, 1983) and Single Strand Conformational Polymorphism (SSCP; Orita *et al.*, 1989; Plomion *et al.*,

1999). However, one of the most promising approaches to EST mapping is through interrogation of Single Nucleotide Polymorphisms (SNPs) (Section 1.3.2.4.3).

1.3.2.4.3 Single Nucleotide Polymorphisms (SNPs)

SNPs, as well as small INserts and DELetions (INDELs), comprise the largest set of sequence variants in the majority of organisms (Kwok *et al.*, 1996; Kruglyak, 1997; Cho *et al.*, 1999). As an example, sequence analysis of the human genome has resulted in the identification of over one million polymorphisms of this type (Sachidananam *et al.*, 2001). Consequently, significant investment has been made in recent years towards the development of technologies for efficient SNP discovery and subsequent screening (Kwok, 2000; Gut, 2001; Shi, 2001). Although many of these approaches were developed for human study, there is great potential for exploitation in plant sciences. Hence, SNP analysis has rapidly been incorporated into crop research, for which there is now significant published data (as reviewed by Rafalski, 2002).

The use of SNPs as genetic markers has great potential for a number of applications within genetic research including genome mapping, integration of genetic and physical maps and analysis of haplotype structure and linkage disequilibrium (Rafalski, 2002). SNP analysis may provide a potentially efficient route to EST mapping, in which the rate-limiting step is often the identification of suitable polymorphisms. This approach will become increasingly relevant as sequence information becomes more easily attainable via more accessible DNA sequence methodologies and exploitation of public databases.

1.3.2.5 Molecular marker summary

The impact of molecular markers on genetic studies has been immense in the last decade. The importance of such approaches has led to significant developments regarding both the identification of new sources of exploitable polymorphisms and also corresponding genotype screening technologies. A large array of different molecular marker classes and interrogation methods has resulted. Therefore, when undertaking any genetic study in which molecular markers are to be used, careful consideration must be afforded to the choice of which marker types will best fulfil the project objectives. For example, in a mapping study in which large numbers of markers may

UNIVERSITY OF BRISTOL

be required, the use of multi-locus marker systems such as AFLPs or RAPDs may be preferable. However, while such dominant marker types can rapidly produce large amounts of segregation data, there may be a trade-off with regard to the genotype information provided. Conversely, co-dominant markers such as RFLPs and microsatellites may be more informative in a particular study but require considerably more time to develop and screen. In such cases (as in the current study), the combination of one or more marker type may be beneficial. However, additional factors such as the relative costs associated with each marker system, the robustness each technique, the amount of exploitable sequence data and differences in the level of expertise and laboratory hardware required to implement the various technologies should also be considered.

1.4 The use of molecular markers in crop improvement

In general, crop improvement refers to the enhancement of the genetic constitution of a crop by means of plant breeding. This is most commonly achieved by the creation of new genetic variation through the crossing of diverse or complementary genotypes with the aim of generating offspring with new combinations of favourable alleles that manifest as phenotypes superior to those of either parent – i.e. trangressive segregation. However, conventional breeding programmes are hampered by problems relating to a lack of understanding regarding the genetic basis of traits and inefficient selection procedures. The variety of ways in which molecular markers can be deployed to overcome such hurdles are discussed below.

1.4.1 Investigations into germplasm relationships

Conventional breeding strategies are generally based on exploitation of existing natural variation within elite, wild or exotic germplasm collections. The efficient use of such resources in a breeding programme can be aided by an understanding of the genetic relationships present within the material and is important for several reasons.

Historically, plant breeding programmes have been very successful in achieving the goal of continuous crop improvement. However, a disturbing consequence of such endeavours has been the narrowing of gene pools in elite lines, resulting in increased genetic uniformity across many crop species. This trend may have serious

implications for future gains by conventional breeding. For example, grain yields in maize have consistently improved since the initiation of US breeding programmes in the 1920s. However, the gene pool of elite maize varieties has now become so shallow that the rate of genetic gain for grain yield is predicted to decline (Duvick, 1992). A further problem of genetic uniformity in crops is the increased vulnerability to biotic stresses, such as pests and diseases, and abiotic stresses, such as drought, cold and salinity.

The conservation and informed exploitation of genetic variation is central to the continued success of plant breeding initiatives. Molecular markers can directly contribute to both these objectives.

1.4.1.1 Conservation and assessment of genetic variation within gene pools

Prior to the advent of molecular markers, the study of relationships in crop gene pools was limited to methods based on phenotype, geographic origin, parentage and history. Such approaches possess significant weaknesses, namely their lack of discriminatory power and a paucity of underlying genetic information (Lee, 1998). In contrast, molecular marker 'fingerprinting' techniques can directly interrogate levels of genetic diversity and the relationships present within germplasm resources (Lee, 1995; Karp *et al.*, 1996; Karp *et al.*, 1997; Van Hintum and Van Treuren, 2002). The huge number of publications in the literature describing results in different crops obtained from using molecular markers for germplasm characterisation are testimony to their successful application in this area (see Kresovich and McFerson, 1992; Bretting and Widrlechner, 1995; Lee, 1995; Lee, 1998). The resulting information can be used to build 'core' collections or genebanks that are representative of the biodiversity present in a crop. Such collections serve to conserve naturally occurring genetic variation that is well characterised, in turn, allowing for a more informed selection of parents for crossing within a breeding programme. For example, when exploiting heterosis in the breeding of hybrid cultivars, diversity studies based on molecular markers can be used to predict the yields of crosses between lines from the same germplasm pool (Melchinger *et al.*, 1992).

In practise, gene bank managers face many problems in day to day management for which molecular markers could make a significant contribution (Van Hintum and Van Treuren, 2002), for example: (1) Acquisition of new material - where molecular markers

could provide information to help determine which samples to include; (2) Maintenance of existing material - where molecular markers could help assess the effects of multiplication practices on diversity loss and assist in the identification of duplicates; (3) Characterisation and evaluation of material- where molecular markers could help determine where and how the variation is distributed, both *ex situ* and *in situ*; (4) Utilisation of, and access to, the germplasm - where molecular markers could provide information for structured access (through better characterisation) or provide tools to screen for specific traits.

1.4.1.2 Exploitation of genetic variation in breeding programmes

Germplasm collections containing wild or exotic relatives of elite cultivars may harbour numerous alleles of potential benefit to plant breeding schemes. The incorporation of such alleles into elite germplasm may result in genotypes that outperform original cultivars (Tanksley and McCouch, 1997). However, this procedure is challenging as exotic germplasms often have inferior phenotypes for many traits when compared to elite lines. The transfer of unfavourable alleles in linkage with desirable alleles through 'linkage drag' may hinder the production of superior genotypes. In addition, problems relating to polygenic inheritance may also limit the utilisation of alleles from these sources to major genes, such as those conferring pest and disease resistances (Sorrells and Wilson, 1997). Molecular markers can greatly assist such introgression strategies by allowing both selection for desired alleles, as well as selection against those that may be detrimental.

1.4.2 Linkage mapping

In general terms, linkage mapping involves the assignment of genetic markers to linkage groups that represent the chromosomes of an organism and the subsequent elucidation of marker orders and genetic distances between them. The concept of genetic mapping is not new. The first genetic maps of crop species were constructed as long ago as the 1930s (Emerson *et al.*, 1935; MacArthur, 1934), predating the discovery of DNA as the hereditary material. Early maps were based on morphological markers and, later, isozymes. However, both approaches have several drawbacks that prevented the realisation of the full potential of genetic mapping strategies (Tanksley *et al.*, 1989). Problems relating to non-neutrality and a general

lack of marker loci were overcome with the development of DNA-based marker systems that had the capacity to provide a rich supply of neutral markers in far greater quantities than was previously possible. The first linkage maps constructed using DNA markers were based on RFLP analysis in humans (Botstein *et al.*, 1980; Wymen and White, 1980). Linkage maps of crop genomes soon followed (McCouch *et al.*, 1988; Slocum *et* al., 1990; Burr and Burr, 1991; Gebhardt *et al.*, 1991; Tanksley *et al.*, 1992). The potential of linkage mapping studies were further enhanced by the advent of PCR-based molecular markers, which can have several advantages over hybridisation-based methodologies such as RFLP (see Section 1.3.2.1). Exploitation of these technological advances has resulted in the generation of genetic linkage maps for many important crop species (reviewed in Paterson, 1996; Kumar, 1999).

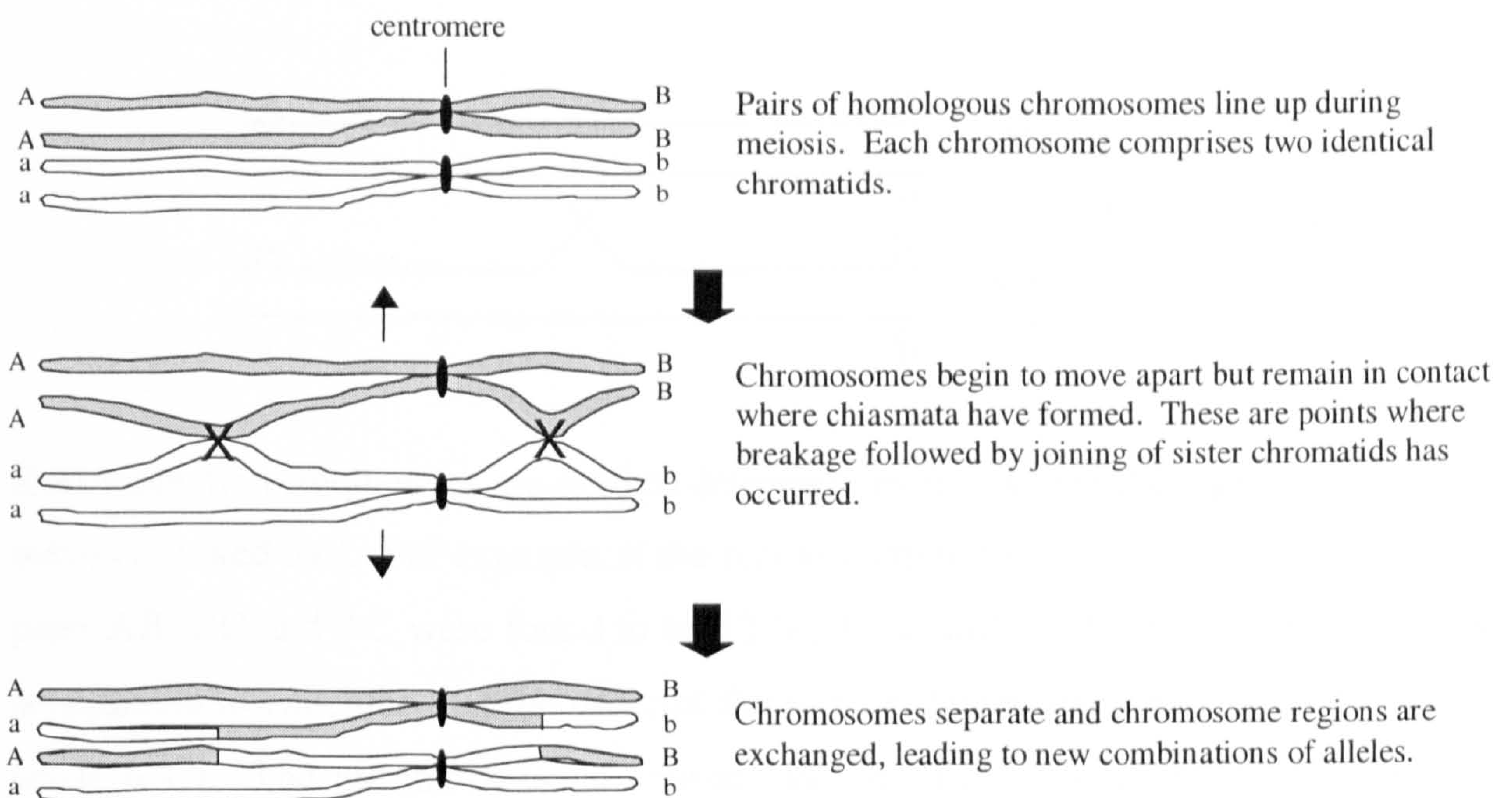
1.4.2.1 The basic underlying principles of linkage mapping

Linkage mapping is based on the genetic principles of segregation and recombination. Segregation is the process by which alleles of a locus separate from one another during the formation of gametes during meiosis. Recombination is the process by which allelic variants of genes are shuffled during gametogenesis, resulting in new combinations of characters in the offspring. It can occur by independent segregation of alleles or by crossing over events between homologous chromosomes.

For example, a diploid individual, heterozygous for a particular locus A (genotype Aa), will give rise to two possible gametes, A or a, which will be produced in a 1:1 ratio. For a second locus B, in the same individual (genotype Bb), a similar outcome would be expected, with equal proportions of gametes of type B or b resulting. If loci A and B are unlinked, i.e. located on separate chromosomes, then the alleles at each locus will undergo independent segregation during meiosis, resulting in the formation of equal proportions of four possible gametes: AB, aB, Ab and ab. If this individual (genotype AaBb) is selfed or crossed to another also of genotype AaBb, the four gametes will recombine at random to form 16 genotypes in the next generation in the following ratio 9A-B-:3AAB-:3A-BB:1aabb (in which the dash refers to either allele being present). Consequently, subsequent matings have the potential to result in progeny with genotypes that differ from the parents at these two loci. Offspring of this type are termed 'recombinants'. However, if locus A and B are linked, i.e. located on

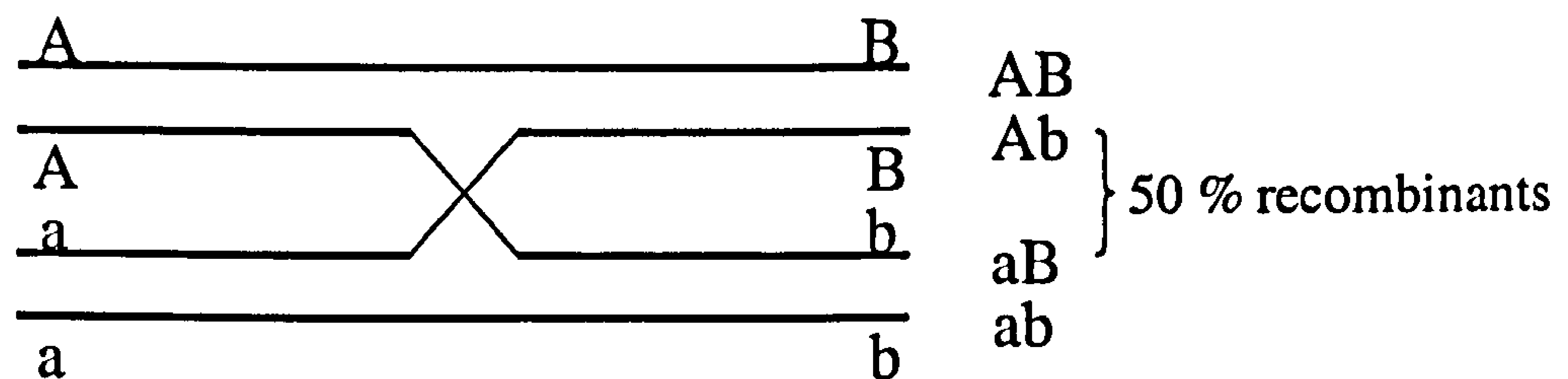
the same chromosome, segregation of alleles will not be independent and recombinants will only arise when crossover events occur between the two loci (Figure 1.8).

Figure 1.8. Schematic representation of recombination by crossover events during meiosis.



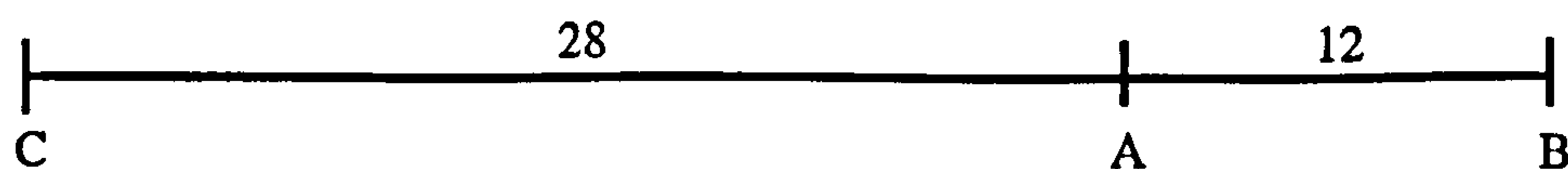
The likelihood of a crossover event occurring between two linked loci is a function of the genetic distance between them, i.e. the chance of a random crossover event occurring between two loci at opposite ends of a chromosome is greater than if the two loci were located near to one another. It is this assumption that underlies the calculation of genetic linkage maps. In a segregating population, it is possible to determine the incidence of recombination events between locus pairs, with the percentage occurrence of such events in the population providing the recombination frequency. The maximum recombination frequency for a linked locus pair is 50 % since only two of the four chromatids involved in meiosis are subject to a crossover event at any one position (Figure 1.9). As this scenario can only occur when loci are located at opposite ends of the chromosome, recombination frequencies are, in general, less than 50 % for most locus pairs.

Figure 1.9. Diagram to show how a maximum of 50 % recombinants can result from a single crossover event involving two chromatids at the four-strand stage of meiosis. Maximum recombination frequencies can only be obtained when two loci are located at opposite ends of a chromosome.



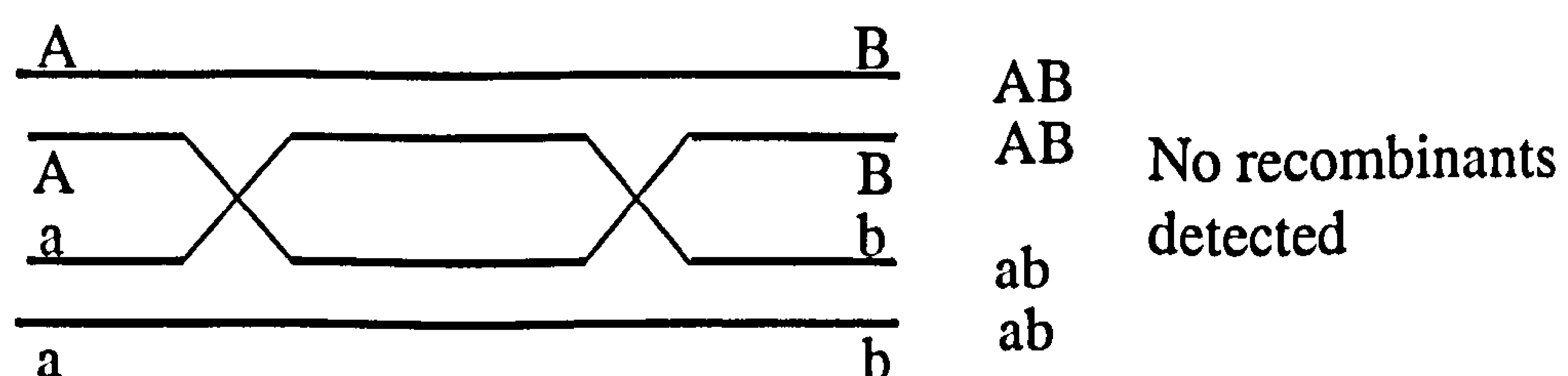
Recombination frequencies are used to determine map orders and the genetic distances between linked loci. For example, if the recombination frequencies for the three locus pairs AB, BC and AC were found to be 12 %, 38 % and 28 %, respectively, it would be possible to conclude that the order of the loci on the chromosome would be C A B (or B A C). The genetic distance between loci A and C would be 28 arbitrary map units, and the distance between loci A and B would be 12 arbitrary map units (Figure 1.10).

Figure 1.10. A simple representation of a hypothetical genetic linkage map based on three loci A, B and C (for explanation, see text).



Underestimation of recombination frequencies, as observed with loci B and C in the example above, may arise as a result of double or multiple crossover events. In such scenarios, although recombination occurs, it may go undetected (Figure 1.11).

Figure 1.11. Diagram to show how double or multiple crossover events can prevent the detection of recombination between linked loci.



For this reason it is important to use mapping functions to convert between recombination frequencies and map distances. The simplest of these was derived by Haldane (1919) but this function assumes that crossovers occur randomly and independently over the entire chromosome. However, this is not quite realistic, as it is generally believed that the occurrence of a crossover event inhibits the incidence of other crossover events in neighbouring regions. This phenomenon is known as 'interference' (Lynch and Walsh, 1998) and is taken into account in a second mapping function derived by Kosambi (1955). These two mapping functions are commonly used today in linkage analysis to calculate map distances from recombination data. Map distances are usually reported in terms of Morgans (M), or more commonly centiMorgans (cM; 1M = 100 cM), after T.H. Morgan who first postulated a chromosomal basis for linkage groups.

As molecular markers have the capacity to generate a vast amount of data points, specialist software packages have been developed to implement linkage analyses, the most commonly used of which are derivatives of MAPMAKER (Lander *et al.*, 1987) and JoinMap™ (Stam and Van Ooijen, 1995; Van Ooijen and Voorips, 2001).

It is important to remember that linkage maps are based on genetic distances calculated from recombination data and, consequently, do not provide a true reflection of the physical distances that may separate loci. The relationship between genetic distance and physical distance may vary enormously between different species as shown in Table 1.3. Furthermore, recombination events do not occur with random distribution across a genome, e.g. recombination rates are suppressed in the vicinity of centromeres (Tanksley *et al.*, 1992). In such regions, linkage maps will, therefore, represent a distortion of the underlying physical distances.

Table 1.3. Approximate relationships between map distance and physical distance in different organisms.

Species	Haploid genome size (kbp)	Map length (cM)	kb per cM
Yeast	2.2×10^4	3700	6
Arabidopsis	7.0×10^4	500	140
Drosophila	2.0×10^4	290	700
Tomato	7.2×10^5	1400	510
Human	3.0×10^6	2710	1110
Maize	3.0×10^6	1400	2140

Source: Lynch and Walsh (1998)

1.4.2.2 Linkage mapping in practice

1.4.2.2.1 Mapping populations

A major consideration in any mapping study is the choice of mapping population that will be used as a basis for linkage analysis. Several types of segregating population are obtainable in crop species, the simplest, and most commonly used of which are derived from crosses between two homozygous parents. Resulting F_1 hybrids can employed in a variety of ways to generate a segregating population. They can be selfed or crossed in pairs to produce a segregating F_2 population, or backcrossed to one of the parents to give a segregating backcross population. F_1 hybrids can also used to generate recombinant inbred lines (RILs) (Burr and Burr, 1991) or used to produce double haploid (DH) populations (Kermicle, 1969; Nitzsche and Wenzel, 1977; Choo, 1981) by the regeneration of plants from haploid pollen and treatment with colchicine to restore the diploid condition. All the mapping population types above have various associated advantages or disadvantages (Kumar, 1999) and the choice of which type of population to use should be influenced by the specific questions that are to be addressed in downstream applications (Haley and Anderson, 1997).

For several plant species, the type of mapping population that can be produced will be severely limited by the reproductive biology of that species. For example, populations derived from fully homozygous parents cannot be produced by conventional crossing strategies in outcrossing plants such as many trees and turf, forage and range grass species unless lengthy intermediary steps are taken, such as the production of doubled haploids. As a result, for outcrossing species mapping populations are often full-

sibling families derived from a cross between two heterozygous parents. Linkage analysis of such outbred populations is possible, although it is far more complicated than with populations derived from inbred lines (Maliepaard *et al.*, 1997) (see Section 3.1.2 for further discussion).

1.4.2.2.2 Choice of molecular marker class

Another important consideration when planning a mapping study relates to the choice of the molecular marker system (or systems) that will be used. As discussed in Section 1.3, marker systems vary in many ways, including their relative levels of informativeness, their single- or multi-locus nature, or the ease with which they can be developed and utilised. Different marker classes may, therefore, have significant associated benefits or drawbacks for a particular linkage mapping study. For example, when mapping in a backcross or DH population derived from inbred lines, AFLPs may be the marker system of choice, as they are informative and have the capacity to provide a large number of robust, polymorphic markers in a relatively short timeframe. In contrast, for mapping studies based on outbred populations, in which more than two alleles per locus may be segregating, the dominant nature of AFLPs may limit the amount of information they can provide. For this type of study, co-dominant marker systems, such as microsatellites or RFLP, are more suitable. These two marker systems are sequence-based and, therefore, have an added benefit in that they may be transferable between mapping populations and are useful in comparative mapping studies, unlike markers produced by arbitrary priming in PCR or AFLP. However, the time required for microsatellite marker development and the laborious screening protocols required with RFLP, may limit the efficiency with large numbers of markers can be mapped by these methods.

1.4.2.3 Applications of linkage maps in crop improvement

Genetic linkage maps are fundamental resources that underpin crop improvement strategies based on molecular genetic approaches. Their primary role is in the genetic mapping of loci underlying traits of agronomic importance (Section 1.4.3) which, when achieved, may provide a plant breeder with key information regarding the genetic basis of such traits. Furthermore, the identification of genomic regions involved in traits of interest serves as a basis for several downstream practical applications, such as marker-assisted selection (MAS) (Section 1.4.5), trait

introgression (Section 1.4.1.2), map-based cloning (Section 1.4.6), candidate gene analysis (Section 1.4.7) and comparative mapping (Section 1.4.4).

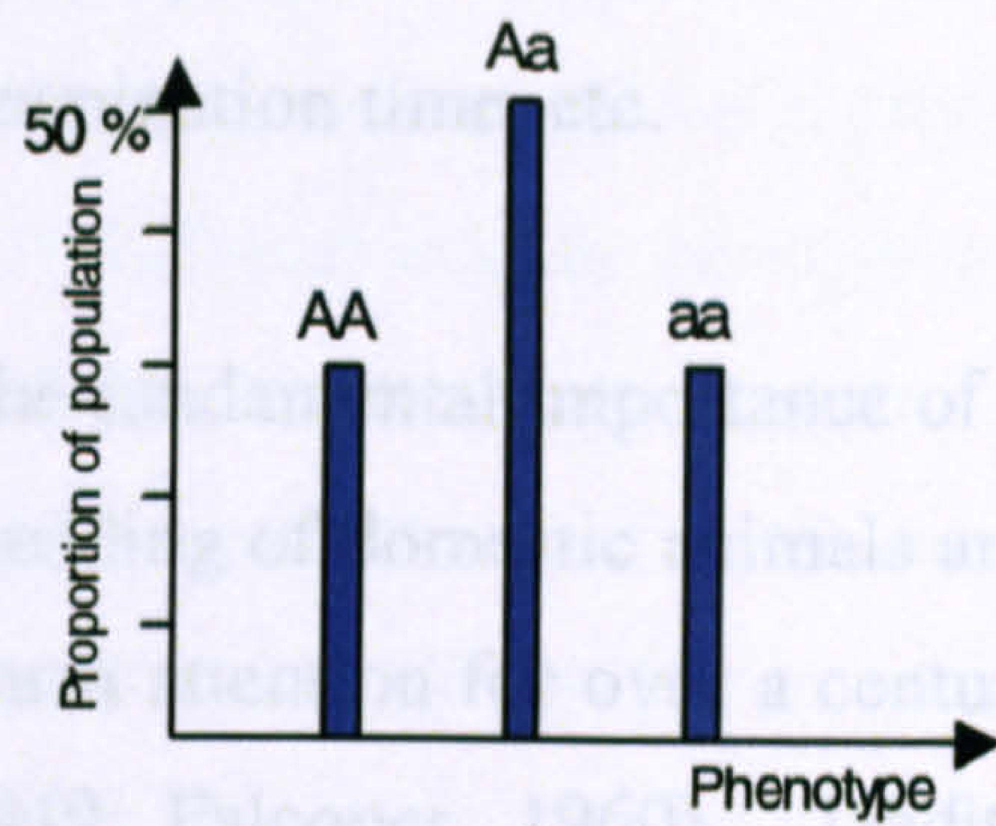
1.4.3 Genetic mapping of important agronomic traits

From a breeding perspective, agronomic traits fall into two categories: qualitative traits and quantitative traits (Stam, 1998). Qualitative traits are generally under the control of so-called 'major' genes that are inherited in simple Mendelian fashion and present distinct phenotypes depending on the allelic forms present (Jones *et al.*, 1997). Examples of traits of this class include many monogenic resistances to crop pathogens (see Kumar, 1999).

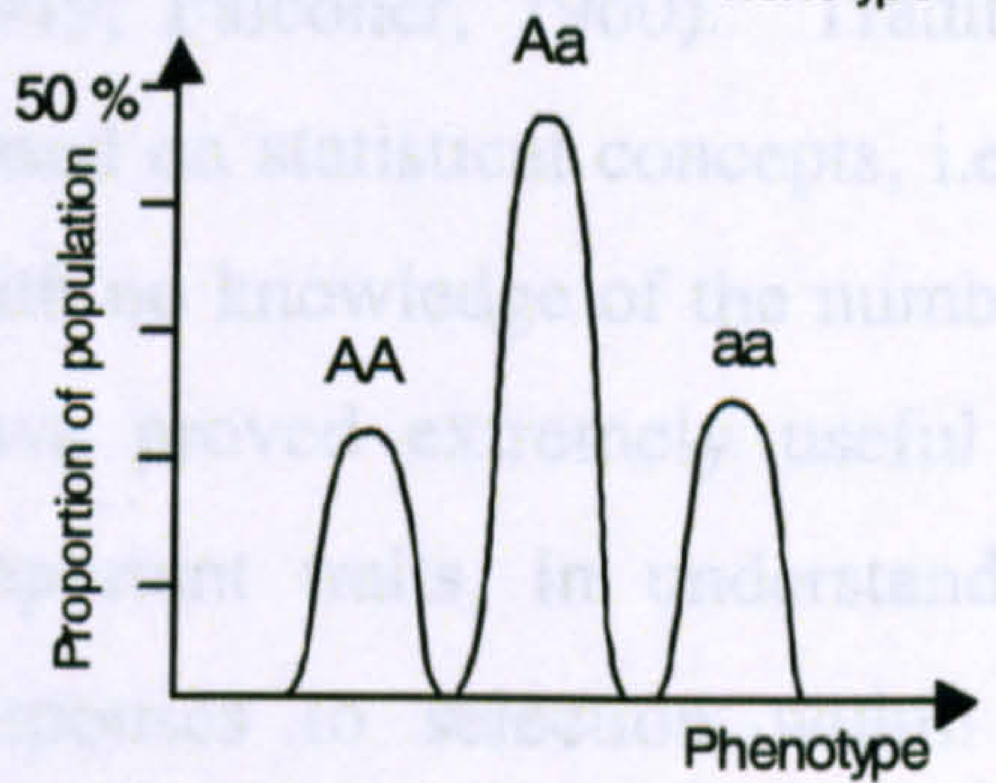
Mapping qualitative traits in segregating populations is relatively straightforward and can be achieved using an approach termed 'Bulked Segregant Analysis' (BSA) (Michelmore *et al.*, 1991). Here, two pooled (bulked) DNA samples are prepared from individual progeny of a segregating population. Each bulk comprises DNA from individuals that are identical for the trait of interest. Two pools that comprise contrasting individuals for the trait of interest are then compared using molecular markers (often AFLPs). Markers that are polymorphic between the two bulked samples will be genetically linked to the gene influencing the trait of interest and can easily be mapped, in turn highlighting the approximate genomic location of the gene of interest. This approach can also be used to identify markers linked to particular map regions in which marker saturation is poor or higher saturation levels are required. Furthermore, this technique provides a rapid means by which markers linked to traits can be identified in populations other than that used to make the genetic map.

However, the majority of agronomic traits are quantitative in nature (Tanksley *et al.*, 1989) being under the control of several genes, so-called 'polygenes' (Mather, 1949). For traits of this type, the observable phenotype is a consequence of both genotypic and environmental variation. As a result, within a population, a quantitative trait will be observable as a continuous range of variation that is more or less normally distributed (Kearsey, 1998) (Figure 1.12). Quantitative variation is thought to arise from the combined effects of allelic differences in structural and regulatory genes which may alter the genes action slightly, resulting in minor phenotypic effects (Kearsey and Pooni, 1996). Some of the genes underlying a quantitative trait may,

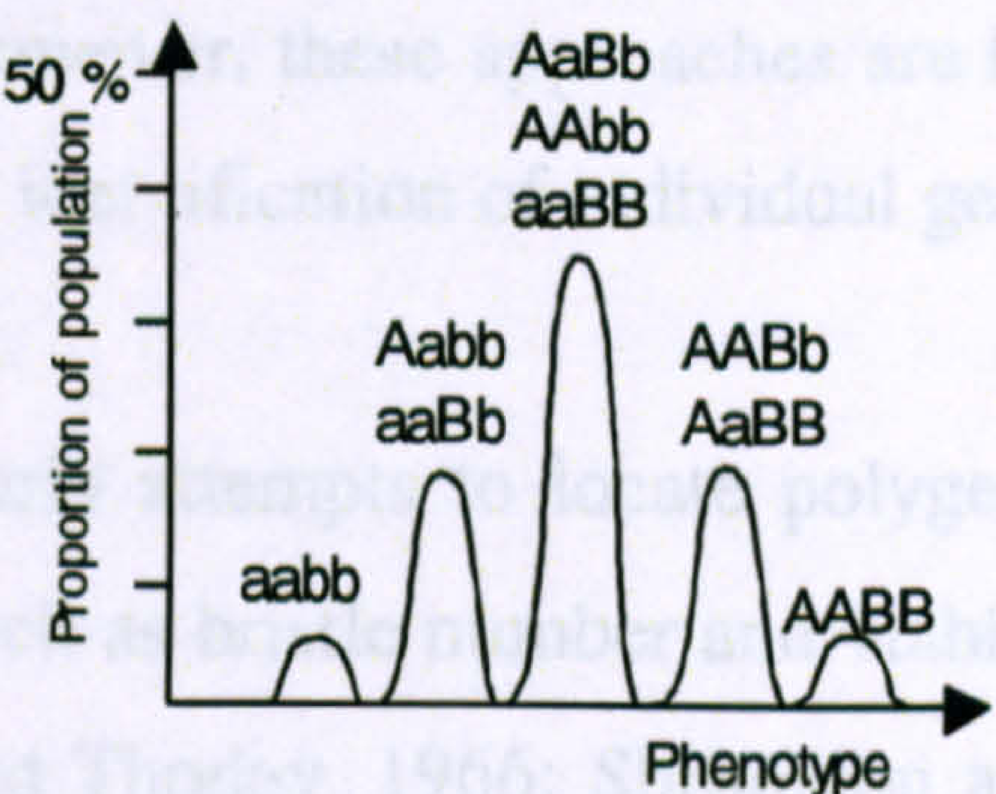
Figure 1.12. Conceptual models for quantitative inheritance. Graphs represent the F2 progeny derived from the selfing of a F1 generation derived from a cross between homozygous parents.



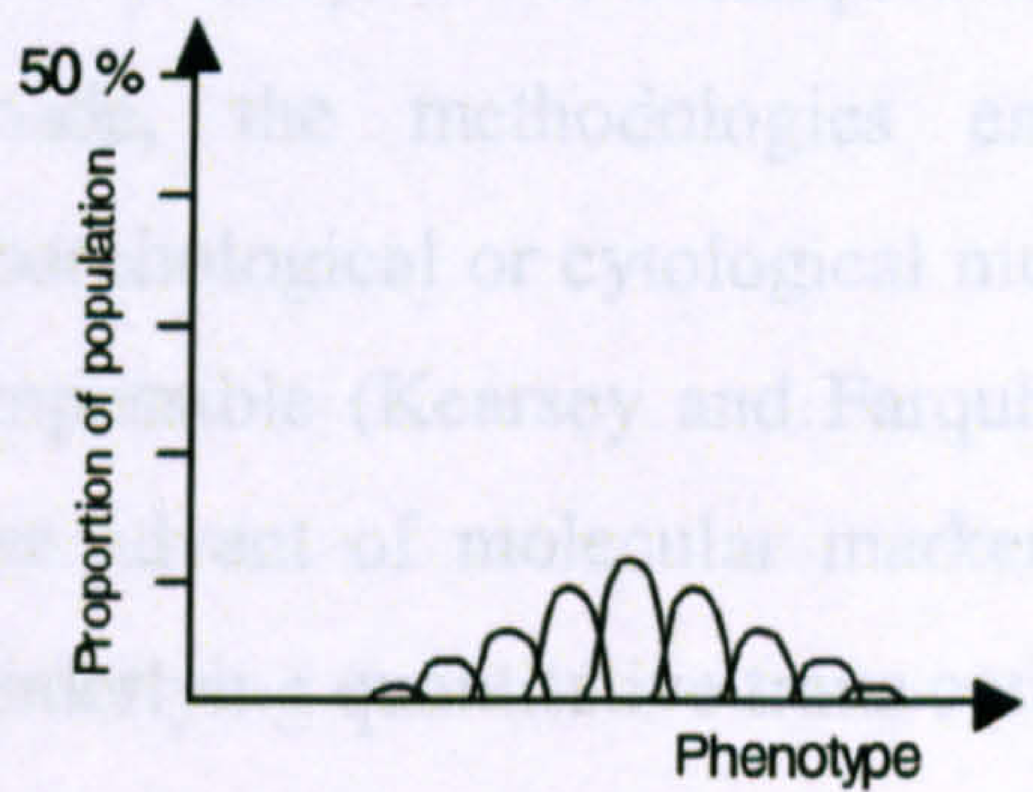
1. Trait under the control of a single locus. No influence of non-genetic factors.



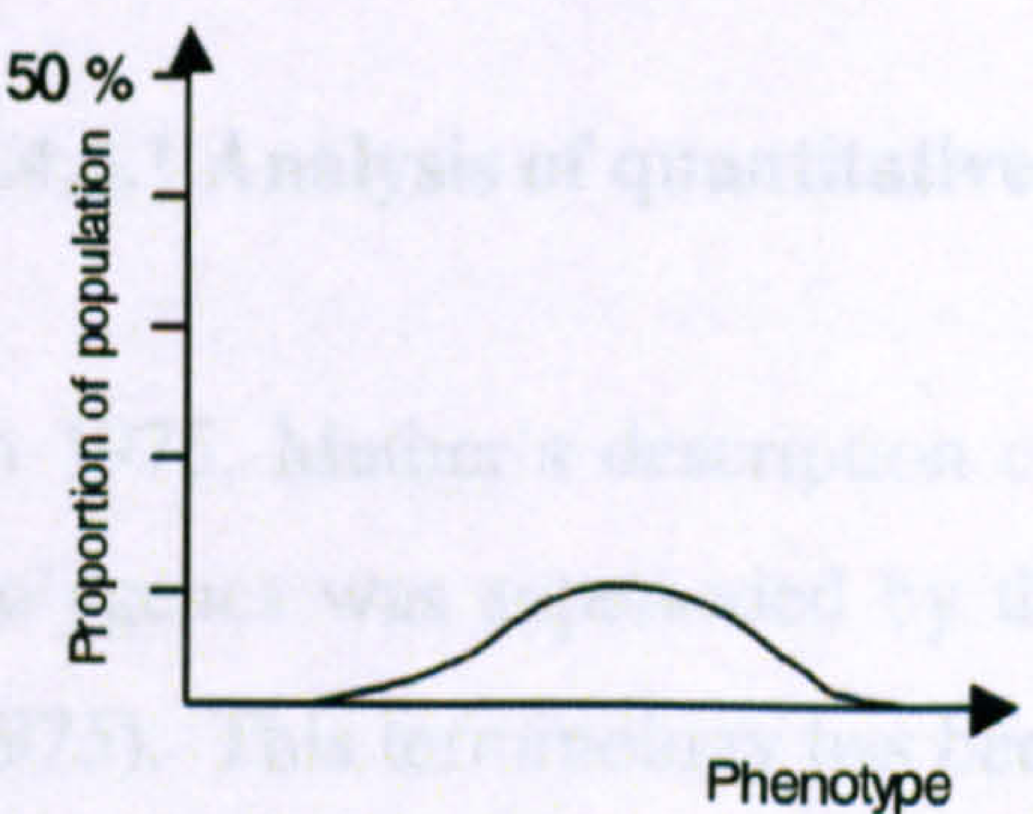
2. Trait under the control of a single locus, but influenced by non-genetic factors.



3. Trait under the control of two unlinked loci (*A* and *B*), and influenced by non-genetic factors.



4. Trait controlled by four unlinked loci, and influenced by non-genetic factors.



5. Trait controlled by many unlinked loci, and influenced by non-genetic factors.

Source: Paterson, 1998.

however, be of qualitative nature, although their effect on overall phenotype will be masked by other sources of variation. An example of a quantitative trait is provided by yield in cereals, which may involve many genes influencing several factors, such as photosynthesis and metabolism, grain and tiller number, root development, germination time, etc.

The fundamental importance of quantitative variation in commercial agriculture, in the breeding of domestic animals and in medicine, has meant that this subject has received much attention for over a century (Galton, 1889; Fisher, 1918; Wright, 1934; Mather, 1949; Falconer, 1960). Traditional methods for analysing quantitative traits were based on statistical concepts, i.e. means, variances and correlations between relatives, with no knowledge of the number and location of underlying genes. Such approaches have proved extremely useful in advancing knowledge regarding the genetics of important traits, in understanding processes such as heterosis, and in predicting responses to selection within breeding schemes (Kearsey and Farquhar, 1998). However, these approaches are limited in application as they do not allow the location or identification of individual genes underlying a quantitative trait (Stam, 1998).

Early attempts to locate polygenes using major gene markers focused on characters such as bristle number and viability in *Drosophila* (Breese and Mather, 1957; Spickett and Thoday, 1966; Shrimpton and Robertson, 1988). While significant progress was made, the methodologies employed were laborious and centred on major morphological or cytological mutants, making the study of populations difficult if not impossible (Kearsey and Farquhar, 1998). These shortcomings were overcome with the advent of molecular markers that provided the means by which individual loci underlying quantitative traits could be studied.

1.4.3.1 Analysis of quantitative traits using molecular markers: QTL analysis

In 1975, Mather's description of the genetic factors underlying quantitative traits as polygenes was superseded by the term Quantitative Trait Loci or QTL (Gelderman, 1975). This terminology has been widely adopted to describe the study of quantitative traits through exploitation of molecular marker technologies, now commonly referred to as QTL analysis.

QTL analysis is made possible by the ability of molecular markers to detect large numbers of neutral polymorphisms in segregating populations. Whilst the majority of polymorphisms detected will be in non-coding regions of the genome, with no direct effect on phenotype, some at least will be linked to QTL (Kearsey, 1998). In such scenarios, the molecular marker locus and the QTL will not segregate independently, and different marker genotypes will be associated with different trait phenotypes. For example, in an F_2 population, a given marker locus, M, may be segregating for two alleles, M and m. If the homozygous condition, MM, was found to be significantly associated with a different phenotype value to individuals of genotype mm, then it would be possible to infer the presence of a QTL influencing the phenotype linked to marker M.

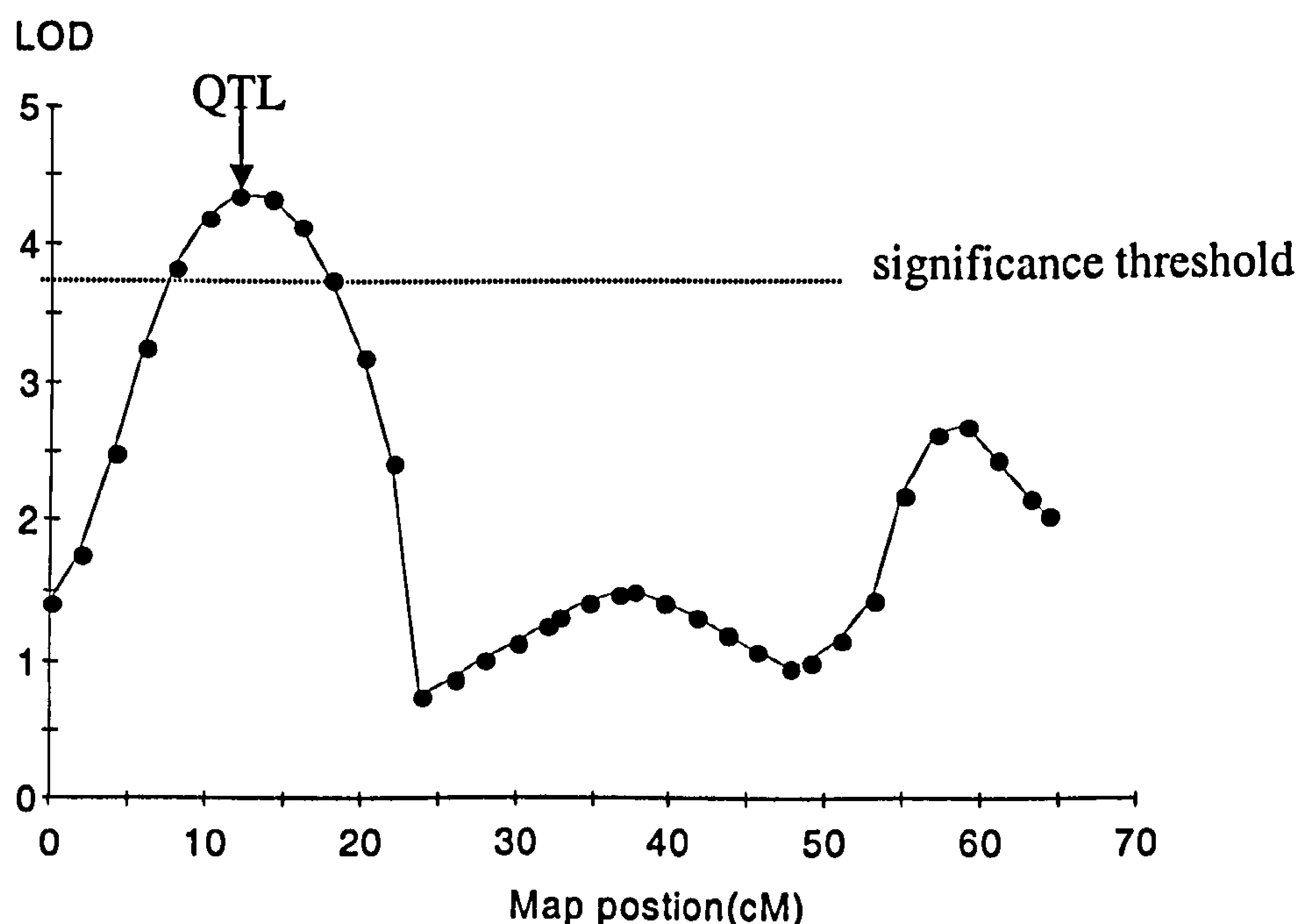
1.4.3.1.1 QTL mapping methodologies

Early QTL analysis focused simply on the examination of associations between individual marker genotypes and phenotypic values, using statistical methodologies, such as the *t*-test or ANOVA, to test the significance of differences in marker genotype for a given trait. Alternative approaches, based on regression of trait values onto marker genotype have also been employed (Kearsey, 1998). These strategies do not draw directly upon information provided by a linkage map and cannot be used to locate a QTL or determine the magnitude of its effect – all that can be concluded is that a marker is linked to a QTL. These types of analysis still have relevance to modern QTL mapping studies as they provide means by which markers can be scanned for QTL associations prior to the availability of a linkage map for the species under study. Furthermore, related approaches may be more applicable to non-parametric mapping studies involving ordinal phenotypic data sets (Van Ooijen *et al.*, 1993).

A more sophisticated approach to QTL analysis, termed 'interval mapping', was proposed by Lander and Botstein (1989). This method exploits information provided by a linkage map and facilitates the calculation of the most likely position of a QTL as well as the magnitude of its effect. Trait information for all adjacent marker pairs is used to determine the likelihood of a QTL being positioned in the interval between them. A likelihood ratio, the LOD (logarithm of odds) score (Barnard, 1949) is then determined at defined points along the chromosome. At the same time, the genetic effects of the QTL and the residual variance are calculated.

A QTL is detected on a chromosome if the LOD score exceeds a predefined threshold, with the most likely position of the QTL corresponding to the highest peak in the LOD profile (Figure 1.13). The decision as to which LOD significance threshold to use in QTL analysis will be dependent on a number of variables, including population type and genome size (Van Ooijen, 1999). This question has received much attention and several articles addressing this issue have now been published (Lander and Botstein, 1989; Van Ooijen, 1992; Churchill and Doerge, 1994; Jansen, 1994; Rebaï *et al.*, 1994; Kruglyak and Lander, 1995; Lander and Kruglyak, 1995; Doerge and Rebaï, 1996; Van Ooijen, 1999).

Figure 1.13. A LOD profile identifying a theoretical QTL by interval mapping. The most likely position of the QTL corresponds to the highest peak of the LOD profile and is located at the 12 cM position. The dotted horizontal line represents the LOD significance threshold, below which, the LOD score is statistically insignificant.



Interval mapping remains the most commonly used and widely accepted methodology for QTL analysis in plants (Kearsey and Farquhar, 1998) although several alternative approaches have been suggested, including methods based on multiple regression (Haley and Knott, 1992) or marker regression (Kearsey and Hyne, 1994). However, interval mapping, in this simple form, only looks for the presence of a single QTL on a chromosome. This can generate misleading results if there are multiple linked QTL present (Haley and Anderson, 1997). In response, more precise methods based on multiple-QTL models, known as MQM mapping or composite interval mapping (CIM), have been developed (Jansen, 1993, 1994; Jansen and Stam, 1994). Here,

following an initial scan to identify QTL, the effect of a particular QTL is then removed from the residual variation, increasing the power and precision of future analyses.

As for linkage analysis, several software packages are available for performing a variety of QTL-related analyses. These include, MapQTL® 4.0 (Van Ooijen *et al.*, 2002), QTL-cartographer (Basten *et al*, 2002), MapMaker/QTL (Lincoln *et al.*, 1992), and Map Manager QTs (Manly and Olson, 1999).

To summarise, QTL analysis is a broad term describing a range of statistical methods that can be used to map loci underlying quantitative traits. This subject has received much scientific attention since the development of molecular marker techniques, both in terms of the underlying statistical theory and its application to a variety of genetic studies. The vast number of publications already available, regarding both these aspects, continues to rise as statistical methods improve and increasing numbers of QTLs are mapped for a variety of traits in growing numbers of species.

1.4.3.1.2 Applications of QTL analysis in crop improvement

The identification and mapping of QTLs can benefit genetic studies in a number of ways. First, QTL mapping can provide information regarding the fundamental genetic basis of a quantitative trait (Stam, 1998). For example, if a population under study segregates for a number of different traits, underlying QTLs can be mapped for each trait separately. This provides a powerful method for determining whether related traits are under the control of distinct sets of QTLs and is exemplified by a QTL study examining yield traits in barley (Stam *et al.*, 1997). Here, a QTL for kernel weight was shown to co-locate with one for total grain yield, suggesting (but not proving) that a single QTL, not unexpectedly, affected both kernel weight and total grain yield. However, a second yield component, number of kernels per ear, was shown to be under the control of a distinct QTL that mapped to a different location to the QTL for total grain yield and kernel weight. This finding indicated that, in the mapping population studied, the number of kernels per ear was unrelated to total grain yield. Hence, information on the causal relationships between correlated traits may be gained by this type of analysis.

A second application of QTL analysis is in the study of genotype-by-environment interactions, i.e. if a population is grown in a range of environments and subjected to QTL analysis, it is possible to distinguish between those QTL that are commonly expressed across all environments and those that are environment-specific. An example is provided by flowering time in *Arabidopsis thaliana,* which was studied under various day length and/or vernalisation regimes (Jansen *et al.*, 1995). Here, several QTL that were commonly expressed over all environments and treatments were identified. In contrast, expression for a number of QTL was shown to be specific to a particular environment or treatment. This type of study may, therefore, be of great importance when breeding crops for a particular environment.

QTL analysis also has the capacity to play an important role in several other aspects of plant breeding and crop improvement, including marker-assisted selections (Section 1.4.5), germplasm enhancement (Section 1.4.1.2), map-based cloning (Section 1.4.6), candidate gene analysis (Section 1.4.7) and comparative mapping (Section 1.4.4).

1.4.4 Comparative mapping

During the mid to late 1980s, as increasing numbers of RFLP-based linkage maps were being generated, it became apparent that the genomes of certain related crop species shared significant organisational relationships (Gale and Devos, 1998a). Early studies highlighted significant similarities between linkage maps of tomato and potato (Bonierbale, 1988) and between the three diploid genomes of hexaploid wheat where the order of genes was shown to be nearly identical (Chao *et al.*, 1988). Such comparisons of genome organisation were made possible by the transferability of RFLP probes between species (Gale and Devos, 1998b), a phenomenon that would later be exploited in mapping studies for species where there remained a paucity of available probes, e.g. probes developed in maize were used to map sorghum (Melake-Berhan *et al.*, 1993) and wheat probes were used to map rye (Devos *et al.*, 1993). Since these early studies comparative genetics has continued to progress, resulting in a greater understanding of the genome relationships between growing numbers of species, both in terms of synteny (conserved clustering of markers/unique sequences) and collinearity (conserved orders of markers/genes). An abbreviated list of examples includes comparative studies between almost all of the economically important grass crops (Devos and Gale, 1997; Gale and Devos, 1998b), between *Arabidopsis* and

Brassica crops (Kowalski *et al.*, 1994; Lagercrantz *et al.*, 1996), between the *Solanaceae* (Livingstone *et al.*, 1999) and between *Arabidopsis* and its close relative, *Capsella rubella* (Acarkan *et al.*, 2000). A general observation has resulted from such studies, in that gene orders, but not intergenic sequences, tend to be conserved within plant families (Gale and Devos, 1998a), with the level of synteny and collinearity present appearing to be related to the evolutionary distance between species (King, 2002). This is a broad statement and exceptions to this paradigm are becoming apparent, particularly at the fine mapping level (Schmidt, 2002).

Comparative genetic studies offer valuable insight into the evolution of genomes (see Gale and Devos, 1998a; Paterson, 2000; King, 2002; Schmidt, 2002 for further discussion) and may impact upon taxonomic thinking (Gale and Devos, 1998b). Furthermore, they also have important applications within crop improvement. It has been demonstrated that different plant species contain homologous genes with very similar functions (Lagercrantz *et al.*, 1996). If sufficient collinearity is present between species under comparative study, then the isolation of a given gene in one species can act as a predictor for the location and subsequent isolation of a corresponding gene in another, e.g. identification of the *Arabidopsis* Gibberellin Insensitive (*GAI*) gene has led to the isolation of orthologous *Rht1* dwarfing genes in wheat (Peng *et al.*, 1997) and dwarfing genes D8 and D9 in maize (Peng *et al.*, 1999). Such approaches are of particular importance in several economically important crops that possess large, complex genomes, such as wheat, maize, oilseed rape, etc. The isolation of genes that have been precisely mapped in complex genomes may be more easily achieved via map-based cloning strategies (Section 1.4.6) in the smaller genomes of model species such as *Arabidopsis* and rice (Gale and Devos, 1998a; Gale *et al.*, 2002). Furthermore, the identification of homologous genes between different species may provide a novel source of allelic variation that has the potential for exploitation in the development of improved crop varieties via transformation technology (Gale and Devos, 1998a).

Comparative studies have recently been widened to encompass QTL. An example is provided by Gale *et al.* (2002) in which map-based comparisons of QTL and major gene loci related to dormancy were made between wheat, maize and rice. Several interesting observations resulted: a set of QTLs identified (in three separate previous studies by Anderson *et al.*, 1993; Sorrells and Anderson, 1995; Bassoi, 2002) on group

1S in wheat were shown correspond to a major rice QTL on the syntenous short arm of rice chromosome 5; a rice QTL on chromosome 3 and a maize candidate gene (see section 1.4.7), *Lw1*, were shown to co-map with the *Phs* gene (Flintham, 2000) and several QTLs on wheat group 4S; a set of wheat QTLs on group 4S co-mapped with a set of rice QTLs on rice group 3S and the *Vp5* gene on maize chromosome 4; wheat QTLs on group 7L in the region of *a-Amy-2* co-mapped with *Il6* in maize and a set of QTLs on rice chromosome 8S. Additional examples of comparative QTL analysis are those between *Brassica* and *Arabidopsis* (Lan and Paterson, 2000; Lan and Paterson, 2001), and between maize, sorghum and foxtail millet (Devos and Gale, 2000).

Comparative genetic approaches clearly have an important role to play in future crop improvement efforts. The ability to exchange genetic information regarding markers, genes and traits between well-studied organisms, such as model species and the major food-crops, harbours great potential for increasing the efficiency with which advances in scientific understanding and crop improvement can be made. This is equally important in the development of novel or lesser-studied crops, where there may be potential for exploiting existing genetic resources from related species. Furthermore, the sequencing of the entire genomes of an increasing number of model plant species, such as *Arabidopsis*, rice and poplar, will provide invaluable resources for future comparative genetic studies.

1.4.5 Marker-assisted selection

Marker-assisted selection (MAS) is the term used to describe the implementation of molecular markers to increase selection efficiency within breeding programmes, by providing a means of selection based on genotype rather than phenotype. Sax (1923) and Thoday (1961) suggested that the presence of a particular gene may be inferred by the presence of a genetic marker that is tightly linked to it, i.e. a gene controlling a trait of interest and a genetic marker are likely to be jointly transmitted to a progeny if they are tightly linked. As the genetic distance between the marker and the gene affecting the trait increases, then the likelihood that they will be transmitted together diminishes, due to the increased chance of recombination events occurring between them. If sufficient linkage is present, then efficient selections based on the marker genotype may be possible without any knowledge regarding the gene itself.

UNIVERSITY OF BRISTOL

Such selection strategies could theoretically provide several opportunities for increasing efficiency within breeding programmes, including increasing the efficiency of trait introgression strategies (Section 1.4.1) and the ability to select for individual factors underlying quantitative traits (QTL), both of which are time-consuming, and in the latter case, difficult to implement by conventional selection means.

The practical realisation of such approaches remained largely unattainable prior to the development of molecular markers, which facilitated the initial identification of marker-trait associations for both qualitative and quantitative traits, and provided a means by which the inheritance of favourable alleles could easily be monitored and tracked through a breeding programme. The ability to select genotypes on the basis of tiny amounts of DNA obtained at the seedling stage, rather than on phenotypes that may take long periods of time to manifest, was also considered a great advantage.

The concept of MAS was initially greeted with great enthusiasm by the plant breeding community. However, it has since become apparent that such approaches may not be as simple to implement as had first been thought. Those traits to which MAS can be most successfully applied are often simple and qualitative in nature, and are often highly heritable. Given the costs associated with molecular marker screening, MAS may be of little benefit in such cases if the trait can easily be scored by phenotypic inspection in the field or greenhouse (Young, 1999). However, there are instances where MAS has been of great value in the breeding for such simple traits, as illustrated by breeding for disease resistance, which is often simple and oligogenic in nature (Melchinger, 1990; Young, 1996). In cases where the phenotype is difficult to score in the field or in glasshouse inoculation tests, or when several pathogens need to be screened simultaneously, MAS can be extremely powerful (Kumar, 1999). An example is provided by breeding for resistance to soybean cyst nematode (SCN), a pathogen that causes significant yield loses in soybean (Wrather, 1998). The laborious and time-consuming nature of phenotypic assays for resistance to SCN resulted in a demand for molecular markers linked to SCN resistance for use in MAS (Young, 1999). As a consequence, several microsatellite markers tightly linked to the SCN resistance gene *rhg1* have now been identified (Mudge *et al.*, 1997; Cregan *et al.*, 1999), one of which, Satt309, has been shown to be extremely useful in selecting for SCN resistance. (Young, 1999). This is just a single example - several other illustrations of the power of MAS in selection for simple qualitative traits, such as

many disease resistances, are also published (see Kumar, 1999), clearly demonstrating the effectiveness of MAS in the manipulation of qualitative traits.

The application of MAS to more complex quantitative traits has proven more challenging. This is a result of factors relating to the increased genetic complexity of such traits, including the potentially large number of genes involved in expression of the trait and also epistatic interactions between genes (Ribaut and Hoisington, 1998). The fact that many genes may be involved, with each having a potentially small effect on the phenotype, suggests that several QTLs must be manipulated simultaneously in order to generate significant performance gain. To achieve this goal, accurate QTL mapping based on large populations (Beavis, 1994), precise phenotyping (Ribaut and Hoisington, 1998), and replications over several years and environments (Young, 1999) will be required. Any of these aspects may be difficult, expensive or time-consuming in practice, making comprehensive descriptions of quantitative traits problematic in terms of molecular markers. However, in recent years, the accuracy with which QTL can be mapped has improved (Young, 1999), largely due to improved experimental design and more powerful statistical methods such as composite interval mapping. These advances may have a positive impact on the use of MAS in manipulating QTL within breeding schemes. Furthermore, the development of novel breeding strategies that can enhance the efficiency of both QTL mapping and MAS simultaneously may prove beneficial. An example of such an approach is advanced backcross QTL (AB-QTL) analysis (Bernacchi *et al.*, 1998).

Despite the problems associated the application of MAS to the improvement of quantitative characters via the manipulation of QTLs, several encouraging results are now emerging in the literature. Examples include selection for malting quality traits and yield in barley (Romagosa *et al.*, 1999; Han *et al.*, 1997; Igartua *et al.*, 2000) and for increased seed protein concentration in soybean (Sebolt *et al.*, 2000). Such results help substantiate the promise of MAS in plant breeding schemes, although clearly this subject is one that will require continued scientific attention in order for the full potential of this technology to be realised.

1.4.6 Map-based cloning

Once a major gene or QTL is mapped, then gene isolation via map-based (or positional) cloning may be a viable prospect (Lee, 1995). This strategy has two major requirements which can easily be met for a large number of traits, in that, 1) individuals within the mapping population must possess genetic polymorphisms underlying the trait of interest, and 2) that the gene(s) responsible for these differences are mapped to a chromosomal position adjacent to regions of DNA for which sequence information is available, i.e. DNA markers. One approach to map-based cloning is provided by so-called 'chromosome walking' in which markers nearby to the gene of interest are used to probe genomic libraries and identify a series of overlapping clones and subsequent contig construction. This is often continued until a marker situated on the opposite side of the gene is reached. However, this method has significant drawbacks in that the isolation of overlapping clones and identification of the gene within the resulting contig can be time-consuming. While this approach is generally applicable to organisms with small genomes, its use may be limited in organisms with more complex genomes, e.g. many plants, as vast lengths of genomic DNA must be traversed. Here, an alternative method, termed 'chromosome-landing', may be more appropriate (Tanksley *et al.*, 1995). This concept is based on the isolation of one or more marker(s) within a physical distance of the target gene that is less than the average size of insert in a yeast artificial chromosome (YAC) or bacterial artificial chromosome (BAC) large insert genomic library. The marker(s) can then be used to screen for the clone containing the gene of interest without the need for chromosome walking and its associated disadvantages. An early example of the efficacy of this approach was provided by the isolation of the *Pto* gene conferring resistance to the bacterial pathogen *Pseudomonas syringae* in tomato (Martin *et al.*, 1993). Other examples of successful map-based cloning outcomes are reported in Arondel (1992), Chang *et al.* (1993), Leung *et al.* (1994), Meyer *et al.* (1994) and Mindrinos *et al.* (1994). More recently, map-based cloning strategies have been applied to the isolation of genes underlying QTL in tomato (Frary *et al.*, 2000; Fridman *et al*, 2000)

1.4.7 The candidate gene approach

The candidate gene approach has been proposed as an alternative method for the characterisation and cloning of important genes and QTLs (Byrne and McMullen, 1996). This approach complements more conventional map-based cloning and insertional mutagenesis techniques, the success of which may be limited due to large genome sizes and/or the lack of transposons in the species under study (Pflieger *et al.*, 2001). It is broadly based on the hypothesis that previously sequenced genes of known function (the candidate genes) may correspond to genes underlying either qualitative or quantitative traits under investigation, with molecular polymorphisms within the candidate gene being related to variation in phenotype.

The selection of candidate genes may be made on a 'functional' or 'positional' basis. Functional candidates are chosen from cloned genes known to be involved in particular biochemical pathways that may play a role in the trait under study, e.g. genes involved in lignin biosynthesis may be potential candidates if wood formation is under investigation. Studies based on cDNA library screening, or cDNA and oligonucleotide microarray technologies can also provide a potential source of functional candidates (Mazeyrat *et al.*, 1998; Causse *et al.*, 1999; Etienne *et al.*, 1999; Aharoni *et al.*, 2000; Giovanonni, 2000). Positional candidate genes are selected on the basis of linkage with a locus under study, i.e. all genes or open reading frames (ORFs) that map to a similar location to a trait of interest may be proposed as potential candidates. This approach has been successfully applied in the cloning of major resistance genes (reviewed by Bent, 1996) and genes involved in flowering time in *Arabidopsis* (reviewed by Koornneef *et al.*, 1998). Positional candidates may also be proposed as a result of comparative mapping studies (see Section 1.4.4) as illustrated by research that led to the implication of the *ACE* gene in human hypertension (Julier *et al.*, 1997). This gene, now known to play a key role in human blood-pressure regulation, was initially proposed as a positional candidate gene due to its location in a region of the human genome that was syntenic to a region of the rat genome containing a QTL affecting hypertension (Hilbert *et al.*, 1991).

Once the choice of candidate genes has been made, they are then screened for potential associations with the trait of interest. In plants, two methods can be employed to achieve this. First, the map location of both the candidate gene and the trait can be

compared. Absolute co-segregation (i.e. with no recombinants) of a locus underlying a qualitative trait with a candidate gene can provide a powerful indication that the candidate gene may indeed underly the trait. Such an approach has proven successful in the characterisation of major loci involved in fatty acid and oil production in *Brassica* (Fourmann *et al.*, 1998) and in mature fruit colour in *Capsicum annuum* (Lefebvre *et al.*, 1998). In addition, several examples have been published in which sequences sharing homology to known resistance genes (R-genes), so-called Resistance Gene Analogues (RGAs), have been shown to co-locate with map regions implicated in disease resistance to a wide range of pathogens (Kanazin *et al.*, 1996; Aarts *et al.*, 1998; Gentzbittel *et al.*, 1998; Hämäläinen *et al.*, 1998; Leister *et al.*, 1998; Rivkin *et al.*, 1999).

Co-segregation approaches are also applicable to the characterisation of QTLs. However, the poor precision with which QTLs are often mapped must be considered; a confidence interval associated with a QTL may cover a vast genomic region that may contain several hundreds, to several thousands, of genes. In such cases, fine mapping experiments, along with statistical testing for QTL-candidate gene associations and subsequent validation studies, such as genetic transformation, are required (Pflieger *et al.*, 2001).

The second screening method for associations between candidate genes relies on determining statistical associations between a candidate polymorphism and a phenotype in unrelated lines. For example, in the study previously cited regarding mature fruit colour in *Capsicum* (Lefebvre *et al.*, 1998), a presence/absence polymorphism of the candidate gene was shown to be significantly associated with fruit colour in 31 unrelated pepper lines, strengthening the hypothesis that the candidate gene was indeed the gene responsible for fruit colour. Again, this type of approach is more complicated when dealing with traits of a more complex nature. The detection of a statistical correlation between a molecular polymorphism within a candidate gene and a QTL does not prove a causal relationship; such associations may occur as a result of the existence of linkage disequilibrium between the polymorphism and the candidate gene (Pflieger *et al.*, 2001). Once more, further validation experiments are required in such scenarios to provide evidence of a causal relationship.

Validated candidate genes may positively impact plant breeding efforts by providing a means of MAS based on polymorphisms within genes that directly affect a trait of interest. This is highly desirable when compared to MAS using molecular markers that are only linked to a trait, where the usefulness of a marker will be limited if the linkage between the trait and marker is not extremely tight.

1.5 Specific aims and objectives of this thesis

This thesis describes a programme of work aimed at taking fundamental steps towards future molecular breeding of SRC willow through the exploitation of molecular markers technologies. The project focused on the following five main objectives:

1) *Development of additional willow molecular markers for genetic studies*

The paucity of molecular markers for use in studies of biomass willows should be addressed. Development efforts should focus on the generation of highly-informative microsatellites to supplement the limited number of these markers that are currently available. The AFLP technique will also be use for rapid marker production and will be used to complement genotype data afforded by the more informative microsatellite markers. Furthermore, the potential to exploit marker resources developed in studies of *Populus* should be examined.

2) *Construction of a genetic linkage map of willow based on the K3 mapping population*

Preliminary efforts to construct a first linkage map of willow based on the K3 mapping population were initiated by Dr. J.H.A. Barker and Dr. A. Karp as part of a previous project at Long Ashton Research Station (LARS) funded by the EU (AIR2-CT92-1617). Segregation data resulting from this work should be supplemented in the current project, with particular regard to the development and screening of newly-developed microsatellite and AFLP markers. When sufficient segregation data is obtained, a genetic linkage map of willow should be constructed.

3) *Establishment and assessment of a large mapping population as a basis for QTL analysis and trait dissection.*

Due to the small progeny number available for the K3 mapping population, a large population more suited to QTL analysis should be established. Efforts should be made to select an appropriate cross in order to achieve segregation of important traits of agronomic relevance. Once chosen, the mapping population should be established in a

field trial and used to generate phenotypic assessment data for use in subsequent QTL analysis and to answer more fundamental questions regarding the genetic basis of important traits, e.g. do they have a quantitative or qualitative basis? Furthermore, the trait assessment data should be used to further understanding of the nature of yield in terms of identifying which yield components (e.g. stem height, stem diameter, number of shoots per stool etc.) are important and may warrant further characterisation.

Primary target traits will include:

- Resistance to *Melampsora* rust
- Resistance to herbivory by Chrysomelid leaf-feeding willow beetles
- Factors that influence yield, i.e. yield components and overall yield

4) *Generation of a second linkage map based on the newly established large-scale mapping population.*

A second linkage map will be constructed based on the newly-established, large mapping population as a basis for subsequent QTL analyses. Focus should be afforded to mapping microsatellite markers due to their highly informative nature in QTL analysis in outbreeding populations. Due to the vast amount of genotyping that will be required, the use of high-throughput screening protocols should be investigated. If necessary, AFLP markers will also be used to improve map coverage.

5) *QTL analysis*

Drawing on information generated in objectives 3 and 4, QTL analysis should be performed in order to identify genomic regions linked to the target traits. Both the newly-established population and the smaller K3 mapping population will be included in these analyses.

1.6 Chapter summary

- Willows are highly heterogenous and comprise over 300 species and numerous hybrids with variation in size, growth form and ploidy levels.
- Several members of the subgenus *Caprisalix* have been highlighted as potentially useful as biomass energy crops when grown as SRC.
- Biomass has the potential to make a significant contribution to renewable energy production, which is beneficial to both sustainable agriculture and the environment.
- SRC willow is a low-cash crop and economic viability is reliant on consistently high yields which can be threatened by pest and disease pressures.
- Breeding of willows specifically for this end-use has resulted in significant progress in recent years, both in terms of increasing yields and resistances, although there is still huge scope for improvement.
- Breeding efforts are hampered by a paucity of information regarding the genetic basis of important traits – a bottleneck that may be addressed through the use of molecular markers.
- Molecular markers, of which there are many classes, can impact crop improvement in several ways and may allow for dissection of a given trait in terms of its underlying factors.
- The identification of molecular markers linked to genes underlying traits of interest may benefit breeding programmes by facilitating marker-assisted selections, by increasing the efficiency of trait introgressions, or by providing a route to gene isolation aimed at direct selection of favourable genes or genetic transformation strategies.
- This thesis describes a programme of work aimed at taking the fundamental steps towards future molecular breeding of SRC willow through the exploitation of molecular markers technologies in linkage analysis and QTL mapping.
- This project aims to identify genomic regions underlying the three main important agronomic traits – resistance to *Melampsora* rust disease, resistance to herbivory by Chrysomelid leaf-feeding willow beetles and finally, yield (and components of yield).

Chapter 2. General materials and methods.

2.1	**BIOLOGICAL MATERIALS**	**59**
2.1.1	Plant material	59
2.1.2	Willow beetles	59
2.2	**NON-BIOLOGICAL MATERIALS**	**59**
2.2.1	Chemicals	59
2.2.2	Radiochemicals	59
2.2.3	Restriction and modifying enzymes	60
2.2.4	Oligonucleotides	60
2.3	**STANDARD SOLUTIONS**	**60**
2.3.1	Molecular biology solutions	60
2.3.2	Liquid bacterial growth medium	61
2.4	**GENERAL MOLECULAR BIOLOGY METHODS**	**61**
2.4.1	Isolation of total plant genomic DNA from leaf tissue	61
2.4.2	Agarose gel electrophoresis	62
2.4.3	Preparation of PCR products for direct sequencing	62
2.4.4	Automated DNA sequencing	63
2.4.5	Sequence analysis	63
2.5	**MOLECULAR MARKER METHODS**	**63**
2.5.1	Microsatellite methods	63
	2.5.1.1 Characterisation of microsatellite library inserts	63

2.5.1.2 Isolation of bacterial plasmid DNA and sequencing of inserts 64
2.5.1.3 Microsatellite primer design 64
2.5.1.4 Microsatellite amplification and detection 65
2.5.1.4.1 Microsatellite amplification and detection : radioisotope method 65
2.5.1.4.2 Microsatellite amplification and detection: [F]dCTP method 66
2.5.1.4.3 Microsatellite amplification and detection: fluorescent primers 66

2.5.2 AFLP methods 66
2.5.2.1 AFLP Method 1 (based on Zabeau and Vos, 1993) 66
2.5.2.2 AFLP Method 2 (Vos *et al.*, 1995) 67

2.5.3 AFLP visualisation 68

2.6 TRAIT ASSESSMENTS 69

2.7 STATISTICAL METHODS 69

2.7.1 Linkage analysis 69

2.7.2 QTL analysis 70

2.7.3 Statistical analysis of trait data 70

2.1 Biological materials

2.1.1 Plant material

All plant material used in this study was collected from willows growing at Long Ashton Research Station (LARS). More detailed descriptions regarding the source of material for each particular experiment are provided in the relevant sections later in the thesis.

2.1.2 Willow beetles

Live adult blue willow beetle specimens (*Phratora vulgatissima*) were collected in 25 ml Sterilin tubes from willows growing at LARS as part of the UK National Willow Collection in 2000 or from SRC willows growing in Redditch, UK in 2001. Collections were made no more than one day in advance for use in experiments. Where storage was required, beetles were transferred to a 30 x 30 cm ventilated Perspex container containing a food supply of willow leaves collected from the same source as the beetles.

2.2 Non-biological materials

2.2.1 Chemicals

All chemicals were purchased from BDH chemicals (Poole, UK), Sigma Chemical Co. (Poole, UK), Promega (Southampton, UK) or GibcoBRL (Uxbridge, UK), unless otherwise stated. Bactotryptone and yeast extracts were obtained from Oxoid (Basingstoke, UK).

2.2.2 Radiochemicals

Redivue™ adenosine 5'-[γ-^{33}P]triphosphate ([γ-^{33}P]ATP; specific activity: >74 TBq mmol^{-1}, ~2000 Ci mmol^{-1}, concentration: 10 mCi ml^{-1}) was purchased from Amersham Biosciences (St Albans, UK).

2.2.3 Restriction and modifying enzymes

Restriction and modifying enzymes were purchased together with associated buffers from Amersham Biosciences or New England Biolabs (Hitchin, UK). *Taq* DNA polymerase was purchased from GibcoBRL (Uxbridge, UK), unless otherwise stated.

2.2.4 Oligonucleotides

Oligonucleotides were synthesised and desalted by Sigma-Genosys Ltd., unless otherwise stated. Fluorophore-labelled microsatellite primers were synthesised and desalted by Applied Biosystems (Warrington, UK).

2.3 Standard solutions

2.3.1 Molecular biology solutions

1x TE buffer	10 mM Tris-HCl, pH 8.0 1 mM EDTA, pH 8.0
1x $TE_{0.1}$ buffer	10 mM Tris-HCl, pH 8.0 0.1 mM EDTA, pH 8.0
10x TBE buffer	0.89 M Tris-borate, pH 8.3 0.025 M EDTA
10x PCR buffer	100 mM Tris-HCl, pH 8.5 500 mM KCl 15 mM $MgCl_2$ 0.01% gelatin

2x Formamide loading buffer for polyacrylamide gel electrophoresis

98% Formamide
10mM EDTA, pH 8
0.05% Bromophenol blue
0.05% Xylene cyanol FF

5x Ficoll loading buffer for agarose gel electrophoresis

15% Ficoll
0.05% Bromophenol blue
0.05% Xylene cyanol FF

2.3.2 Liquid bacterial growth medium

Luria-Bertani (LB) medium

Per litre:

10 g tryptone
5 g yeast extract
10 g sodium chloride

The pH was adjusted to 7.5 with solid NaOH and the medium was autoclaved prior to use.

2.4 General molecular biology methods

2.4.1 Isolation of total plant genomic DNA from leaf tissue

Genomic DNA extractions were performed on fresh leaf tissue collected from field sites at LARS. Care was taken not to collect during the annual onset of rust infection. Isolation of DNA from the K3 mapping population was performed using the Nucleon™ Phytopure™ Genomic DNA Extraction Kit (Amersham Biosciences) according to the Manufacturers' instructions, except that 10 mM 2-mercaptoethanol was added to Reagent 1 before use. For this method ~ 125 mg of fresh leaf tissue was used in each extraction High-throughput isolation of DNA from the K8 mapping population was performed using the DNeasy 96 Plant Kit (Qiagen), using 50 mg of fresh leaf starting material, according to the Manufacturers' instructions. All other

DNA extractions were performed using the DNeasy Plant Mini Kit (Qiagen) according to the Manufacturers' instructions.

Following all DNA extraction protocols, the quality and quantity of DNA was assessed by agarose gel electrophoresis (Section 2.4.2) using 0.8% agarose gels. For DNA quantification, samples were loaded alongside known concentrations of uncut lambda DNA (Sigma) and quantified using the volumetric analysis tools in Quantity One (Version 4.2.2) software.

2.4.2 Agarose gel electrophoresis

DNA fragments were size separated by electrophoresis through agarose gels containing between 0.8% and 2% agarose (GibcoBRL), 1x TBE buffer and 5 μg ml^{-1} ethidium bromide. Electrophoresis was performed using Horizon™ 20-25 gel electrophoresis apparatus (GibcoBRL) and Bio-Rad power packs (PAC 300/3000). DNA samples were mixed with 1/5 volume 5x Ficoll loading buffer and loaded into the wells of gels submerged in 1x TBE buffer. DNA fragments were separated at 100 V for 1-2 hours. Following electrophoresis, DNA was visualised using a 302 nm UV transilluminator (Bio-Rad GelDoc™ 2000 Image Analyser). Images were captured using Quantity One (Version 4.2.2) software (Bio-Rad) and printed using a Mitsubishi digital graphic printer. For DNA fragment size comparisons, 123 bp ladder (GibcoBRL) or 1 Kb DNA ladder (GibcoBRL) was loaded alongside the DNA samples.

2.4.3 Preparation of PCR products for direct sequencing

PCR products (25 μl) were separated by electrophoresis on agarose gels and excised using a clean razor blade. DNA was then purified from gel slices according to Hanley *et al.* (2000) or alternatively, the QIAEXII Gel Extraction Kit (Qiagen) was used according to the Manufacturers' instructions. Following the former method, precipitated DNA was re-suspended in 20 μl sterile distilled water (SDW). In the latter method, DNA was eluted in 30 μl SDW. A 3 μl aliquot of the template was routinely used in cycle sequencing reactions (Section 2.4.4).

2.4.4 Automated DNA sequencing

Automated DNA sequencing was performed using either the ABI PRISM® BigDye™ Terminator Cycle Sequencing Kit with an ABI PRISM® 377 DNA Sequencer or with the ABI PRISM® BigDye v3.0 Terminator Cycle Sequencing Kit and an ABI PRISM® 3100 Genetic Analyser (Applied Biosystems), using a 36 cm capillary array with POP™-6 polymer (Applied Biosystems). The Manufacturers' protocols were followed, except that all cycle sequence reactions were performed in 10 µl total volumes as follows:

Template DNA	2 – 5 µl
Primer (100 ng μl^{-1})	1 µl
BigDye Terminator Master Mix	4 µl
SDW	to 10 µl total volume

Unincorporated dye-terminators were removed prior to sample electrophoresis according to the isopropanol precipitation protocol recommended by the Manufacturer.

2.4.5 Sequence analysis

Sequence analysis and editing was performed using Sequencher™ (Gene Codes Corporation, Michigan, USA), Autoassembler® (Applied Biosystems) and Chromas (Griffith University, Queensland, Australia) software packages.

2.5 Molecular marker methods

2.5.1 Microsatellite methods

2.5.1.1 Characterisation of microsatellite library inserts

For characterisation of additional inserts from the microsatellite-enriched, small insert, willow genomic library (Edwards *et al.*, 1996; Section 3.2.5), 5 µl of each library clone (previously stored at -80°C as a glycerol stock) was used to inoculate 4 ml fresh LB liquid medium containing filter-sterilised ampicillin at a final concentration of 100 µg ml^{-1}. Cultures were incubated overnight at 37°C in a shaking incubator at 200 rpm.

2.5.1.2 Isolation of bacterial plasmid DNA and sequencing of inserts

Following overnight culture, bacterial suspensions were transferred to 2 ml screw-top tubes (Alpha Laboratories, Eastlieigh, UK) and the cells pelleted by centrifugation at 8000 rpm for 2 min in a Microcentaur microfuge (Sanyo). Plasmid DNA was then isolated using a QIAprep Spin MiniPrep Kit, according to the Manufacturers' instructions, with a final elution in 50 µl sterile distilled water. DNA quantity and quality were assessed by agarose gel electrophoresis as described in Section 2.4.2.

Library inserts were sequenced as described in Section 2.4.4. Primarily, only the M13 Forward Universal primer (5'- CCCAGTCACGACGTTGTAAAACG - 3') was used in cycle sequencing reactions. However, in some instances where insufficient read length was obtained, the M13 Reverse Universal primer (5'- AGCGGATAACAATTTCACACAGG - 3') was used as an alternative. Routinely, 3 µl plasmid template DNA was included in each sequence reaction. However, this amount was raised or lowered accordingly with individual samples for which low or high plasmid yields were observed by agarose gel electrophoresis.

2.5.1.3 Microsatellite primer design

Following software-based screening for duplications, forward and reverse primers were designed to the flanking regions of unique microsatellites using the Primer Version 0.5 software package (Whitehead Institute for Biomedical Research, Massachusetts). Suitable primers were identified for each locus according to the following criteria: primer length = 21 bp; T_m = 54 – 62°C; %GC = 40-60%; length of PCR product = 75 – 300 bp. Primers that conformed to these criteria were then examined independently for self-complementarity along the primer and also for complementarity specifically at the 3' terminus. Potential forward and reverse primers per locus were then tested for pairwise complementarity and compatibility of annealing temperatures. The most suitable primer sets were then synthesised by Sigma-Genosys Ltd. for testing as described in Tables 3.4 and 5.4.

2.5.1.4 Microsatellite amplification and detection

Three methods for detection of microsatellites were used in this study: 1) radioisotope-based detection using [γ-^{33}P]ATP; 2) detection via incorporation of fluorescently-labelled 2'-deoxycytidine 5'-triphosphate ([F]dCTP); 3) detection via use of fluorescently-labelled primers in PCR.

2.5.1.4.1 Microsatellite amplification and detection : radioisotope method

The forward primer of each microsatellite primer set was end-labelled with [γ-^{33}P]ATP in a 0.5 µl total volume reaction comprising: 1 x One Phor All (OPA) buffer, 25 ng forward primer, 0.5 µCi [γ-^{33}P]ATP, 1 U T4 polynucleotide kinase (PNK)(Pharmacia) and SDW to total volume. The labelling reaction was incubated at 37°C for 30 min, after which the enzyme was heat denatured at 68°C for 10 min. Labelled primer was stored below 4°C until required.

PCR reaction mixtures contained 10 ng template DNA, 25 ng labelled forward primer, 25 ng reverse primer, 200 µM of each dNTP (Promega), 0.5 U *Taq* DNA polymerase (GibcoBRL) and 1 x PCR buffer. Thermocycling conditions were as follows: 35 cycles of 94°C for 40 s, 54°C for 60 s and 72°C for 60 s, followed by a final 72°C extension period of 10 min. A GeneAmp® PCR system 9700 (Applied Biosystems), with 9600 ramping setting, was used for all amplifications. Amplification products were mixed with an equal volume of 2 x formamide loading buffer and denatured at 94°C for 3 min and placed immediately on ice prior to loading.

Amplification products were resolved by denaturing polyacrylamide gel electrophoresis (PAGE) on 6% gels (70 ml volume) comprising 29.4 g urea, 10.5 ml acrylamide/bisacrylamide 19:1 (National Diagnostics) in 1 x TBE buffer. Following the addition of 75 µl TEMED and 375 µl 10% ammonium persulphate (APS), gels were poured and allowed to polymerise for at least 1 h. Gels were then pre-electrophoresed at 55 W for at least 15 min on GibcoBRL Model S2 gel apparatus using a Bio-Rad PAC 3000 power pack. Denatured amplification products were loaded and separated for approximately 2 hours. Gels were then transferred to Whatman 3MM filter paper, covered with Saran Wrap and dried under vacuum at 80°C using a Bio-Rad model 583 gel dryer. Radio-labelled DNA fragments were visualised

by autoradiography, where Biomax MR-1 film (Kodak) was exposed to the gel in an autoradiograph cassette overnight at room temperature. Autoradiographs were developed using an X-Ograph compact X2 automated developer.

2.5.1.4.2 Microsatellite amplification and detection: [F]dCTP method

Fluorescently-labelled dCTP coupled to [R110] or [R6G] rhodamine dyes was purchased from Applied Biosystems and used to label PCR products according to the Manufacturer's instructions. Thermocycling was performed as described in Section 2.5.1.4.1. Electrophoresis of PCR products was performed using 36 cm gels on an ABI PRISM 377 DNA Sequencer with GeneScan-500 [ROX] as the size standard (Applied Biosystems) according to the Manufacturer's instructions. Resulting electropherograms were analysed using Genotyper® software (Applied Biosystems).

2.5.1.4.3 Microsatellite amplification and detection: fluorescent primers

Protocols used for both single locus and multiplex PCR protocols, is described in detail in Section 5.2.7.

2.5.2 AFLP methods

Two methods were used to generate AFLP markers in this study based on two separate techniques as described in Hanley et al. (2000). Both methods were based on the protocols published by Zabeau and Vos (1993) and Vos et al. (1995). Throughout this thesis these methods are referred to as AFLP Method 1 and Method 2, respectively.

2.5.2.1 AFLP Method 1 (based on Zabeau and Vos, 1993)

This technique was performed essentially as described in Zabeau and Vos (1993). In brief, 500 ng genomic DNA was digested with 12.5 U *Mse*I (New England Biolabs) and 12.5 U *Pst*I (Pharmacia) in 1 x OPA buffer (10 mM Tris-acetate, 10 mM magnesium acetate, 50 mM potassium acetate, pH 7.5) at 37°C for 1 hr, in a total volume of 50 µl. Restriction fragments were then ligated to *Mse*I (50 pmol) and biotinylated *Pst*I (5 pmol) adapters in a reaction mixture comprising 1 U T4 DNA

ligase in the presence of 1 mM ATP and 1 x OPA buffer. Adapter sequences were as follows:

*Mse*I adapter: 5'- GACGATGAGTCCTGAG - 3'
3'- TACTCAGGACTCAT - 5'

*Pst*I adapter:: 5'- biotin – CTCGTAGACTGCGTACATGCA - 3'
3'- CATCTGACGCATGT - 5'

Streptavidin-coated magnetic beads (Dynal) (Tong and Smith, 1992) were used to select for biotinylated fragments as described in Mathes *et al.* (1998). These fragments formed the template for PCR using primers designed to anneal to the adapter sequences. Primer sequences were as follows, with n denoting additional selective nucleotides:

*Mse*I primer: 5'- GATGAGTCCTGAGTAA + nnn - 3'
*Pst*I primer: 5'- GACTGCGTACATGCAG + nn - 3'

Details of the selective nucleotides used are provided in Table 3.3. For use in PCR, the *Mse*I primer was end-labelled using [γ-^{33}P]ATP as described in Vos *et al.* (1995). Each PCR reaction was performed in a 20 µl total volume containing: 2 µl template DNA (bound to beads), 1 x PCR buffer, 5 ng labelled *Mse*I primer, 25 ng unlabelled *Mse*I primer, 20 ng *Pst*I primer, 200µM of each dNTP and 1 U *Taq* DNA polymerase (GibcoBRL). Amplification was performed using a GeneAmp® PCR system 9700 (Applied Biosystems), with 9600 ramping setting, with the following parameters: 10 cycles of 94°C for 30 s, 62°C for 30 s, and 72°C for 60 s, followed by 25 cycles of 94°C for 30 s, 56°C for 30 s, and 72°C for 60 s. A final extension step of 10 min at 72°C completed the program.

2.5.2.2 AFLP Method 2 (Vos *et al.*, 1995)

For this approach, 50 ng genomic DNA was digested with 5 U *Eco*RI and 5 U *Mse*I in 1 x OPA buffer in a total volume of 40 µl for 1 h at 37°C. Adapters were then ligated in a reaction mixture containing: 5 pmol *Eco*RI adapter, 50 pmol *Mse*I adapter, 1 x OPA buffer, 1 mM ATP and 1.4 U T4 ligase, and incubated for 3 h at 37°C. The

*Eco*RI adapter sequence is shown below (the *Mse*I adapter sequence was as for AFLP Method 1):

*Eco*RI adapter: 5'- AATTGGTACGCAGTC - 3'
3'- CCATGCGTCAGATGCTC – 5'

Ligated DNA was diluted 10-fold in 1 x $TE_{0.1}$ buffer and 5 µl used in the following 20 µl total volume pre-amplification reaction: 1 x PCR buffer, 200 µM of each dNTP, 50 ng *Mse*I (+1 selective nucleotide) primer, 50 ng *Eco*RI (+1) primer and 0.5 U *Taq* DNA polymerase. Amplification was performed using a GeneAmp® PCR system 9700 (9600 ramping setting) with the following cycling profile: 65°C for 5 min followed by 30 cycles of 94°C for 30 s, 56°C for 30 s and 72°C for 60 s. Amplification products were then diluted 10-fold in 1x $TE_{0.1}$ buffer.

For the selective PCR amplification, 5 ng *Eco*RI (+3) primer (5' – GACTGCGTACCAATTC + nnn – 3') was end-labelled in a total volume of 0.5 µl with 1 x OPA, 0.5 µCi [γ-^{33}P]ATP and 0.05 U T4 PNK at 37°C for 30 min, after which the enzyme was heat denatured at 68°C for 10 min.

Selective amplifications were performed in 10 µl total volumes containing 2.5 µl diluted pre-amplified DNA template, 200 µM of each dNTP, 15 ng *Mse*I (+3) primer, 5 ng labelled *Eco*RI (+3) primer and 0.25 U *Taq* DNA polymerase. A touchdown profile was used in thermocycling as follows: 13 cycles of 94°C for 30 s, 65°C* for 30 s and 72°C for 60 s, followed by 23 cycles of 94°C for 30s, 56°C for 30 s and 72°C for 60 s, where * denotes a reduction of 0.7°C in each cycle. Details of selective nucleotides used are provided in Tables 3.3 and 5.6.

2.5.3 AFLP visualisation

AFLP fragments generated by either method were resolved by denaturing PAGE, as described in section 2.5.1.4.1, and visualized by autoradiography, as described in Section 2.5.1.4.1, except that 4.5% acrylamide gels (75 ml total volume) were used (31 g Urea - 6 M, 9 ml acrylamide/bisacrylamide 19:1 and 1 x TBE buffer) and exposure times of 2–3 days were routinely used.

2.6 Trait assessments

All field- and laboratory-based trait assessments are described in relevant chapters later in the thesis (Sections 4.2.6, 4.2.7 and 6.2.1).

2.7 Statistical methods

2.7.1 Linkage analysis

Linkage analyses were performed using JoinMap® version 3.0 software (Van Ooijen and Voorips, 2001) as this has the capability to perform linkage analysis on outbred progenies involving markers of different segregation types. As for JoinMap 2.0, version 3.0 uses the estimation procedures for cross-pollinators as described in Maliepaard *et al.* (1997). This procedure uses all information for any combination of markers, segregating from two alleles in just one parent, through two alleles in both parents, to three and four segregating alleles in both parents. This software package was purchased from Plant Research International, Wageningen, The Netherlands. For analysis within this software package (and for ease of reference in this thesis), markers were assigned codes depending on their segregation type (Table 2.1).

Table 2.1. Marker segregation codes for use in JoinMap® 3.0.

Code	Segregation type	Possible genotypes	Expected segregation ratio
aaxab	marker heterozygous in first parent only, two alleles	aa, ab	1:1
abxaa	marker heterozygous in second parent only, two alleles	ab, aa	1:1
abxab	marker heterozygous in both parents, two alleles	aa, ab, bb	1:2:1 or 3:1‡
abxac	marker heterozygous in both parents, three alleles	aa, ab, ac, bc	1:1:1:1
abxcd	marker heterozygous in both parents, four alleles	ac, ad, bc, bd	1:1:1:1

‡: dominant markers segregate 3:1 (presence:absence) and heterozygotes cannot be distinguished from homozygotes. Dominant markers of this type are termed *a0xa0*, in accordance with Maliepaard *et al.* (1997), for the purpose of presentation within this thesis. For analysis in JoinMap, an additional code (aa, b-) was included to indicate dominance (Van Ooijen and Voorips, 2001).

Further details regarding the exact parameters and methodologies used for linkage analyses are provided in the relevant chapters. (Section 3.2.11; Section 5.2.11).

2.7.2 QTL analysis

All QTL analyses were performed using MapQTL® version 4.0 software (Plant Research International) as outbred progenies can be analysed in interval mapping (Lander and Botstein, 1989) using an 'all-markers' approach (Maliepaard and Van Ooijen, 1994), where segregation information from all linked markers irrespective of segregation type (with two up to four segregating alleles) can be utilised. A more comprehensive description of the QTL methodologies used is provided in Section 6.2.4.

2.7.3 Statistical analysis of trait data

All additional statistical analyses of data generated in trait assessments were performed using the GenStat statistical package (© Sixth Edition, Lawes Agricultural Trust, Rothamsted Experimental Station, 2002), as described in Sections 6.2.2 and 6.2.3.

Chapter 3. Construction of the K3 linkage map

3.1 INTRODUCTION 73

3.1.1 Genetic mapping in willow: general considerations 73

3.1.2 Linkage analysis in outbreeding species 73

3.2 MATERIALS AND METHODS 76

3.2.1 The K3 mapping population 76

3.2.2 DNA extraction 77

3.2.3 Development of additional AFLP markers 77

3.2.4 Scoring of AFLP markers 77

3.2.5 Development of additional microsatellite markers 77

3.2.6 Microsatellite primer testing and screening 78

3.2.7 Scoring microsatellite markers 79

3.2.8 Poplar microsatellite testing 79

3.2.9 Sequencing of poplar microsatellite products at locus PMGC223 80

3.2.10 ESTP markers 80

3.2.11 Map construction 81

3.3 RESULTS 82

3.3.1 AFLP markers 82

3.3.2 Microsatellite markers: development and screening 84

3.3.3 Poplar microsatellite primers 88

3.3.4 ESTP markers 89

3.3.5 Construction of the parental maps 89

3.3.6 Construction of the K3 consensus map 95

3.4 DISCUSSION 107

3.4.1 Choice of molecular markers 107

3.4.2 AFLP markers 108

3.4.3 Willow microsatellite markers 108

3.4.4 Poplar microsatellites 109

3.4.5 ESTP markers 110

3.4.6 Segregation analysis 111

3.4.7 The K3 parental maps 112

3.4.8 The K3 consensus map 112

3.5 CHAPTER SUMMARY 114

3.1 Introduction

This chapter describes the construction of a first genetic linkage map of biomass willow (*Salix viminalis*) based on microsatellite and AFLP markers. This line of research was initiated by Dr. J.H.A. Barker and Dr. A. Karp as part of a previous project at Long Ashton Research Station (LARS) funded by the EU (AIR2-CT92-1617). At this time (1995), no linkage maps of any *Salix* species had been published. During this previous project, marker segregation data was generated for more than 400 AFLP and 13 microsatellite markers. However, preliminary analysis of the available data indicated that insufficient resolution of linkage groups could be achieved (J.H.A. Barker, personal communication). The development and screening of further markers was therefore undertaken and the existing markers were re-scored and checked. Linkage analysis and map construction were then performed using the amalgamated data set. The results were published in the journal Theoretical and Applied Genetics (Hanley et al., 2002). This chapter provides a fuller account of all the work.

3.1.1 Genetic mapping in willow: general considerations

Genetic linkage mapping in willow presents a formidable challenge given the relatively high haploid chromosome number (19) and the dioecious, outcrossing and highly heterogenous nature of willow species. In addition, at the original onset of this work, molecular marker development in willow was extremely limited, particularly when compared to major crops and significant investment into developing molecular markers was, thus, required

3.1.2 Linkage analysis in outbreeding species

Linkage mapping in outbreeding plant species is often based on full-sib families and is, consequently, far more complicated than mapping studies focusing on progenies derived from two fully homozygous parents, e.g. backcross, F_2, RILs and double haploid populations. The added complexity of such studies results from the number of alleles that may be segregating at a given locus and uncertainties over the linkage phase of the loci (Maliepaard *et al.*, 1997), i.e. in a full-sib family, up to four alleles per locus may be segregating at a given locus and linkage phases are usually unknown.

In contrast, in a mapping population derived from fully homozygous parents, all loci will segregate for a maximum of two alleles with all alleles from the same parent being in coupling phase. Furthermore, the number of segregating alleles detected may vary between marker classes, i.e. multi-allelic markers, such as RFLPs and microsatellites, may detect all alleles at a given locus, whereas AFLP and RAPD systems generally manifest as the presence or absence of a DNA fragment. Also, dominant and co-dominant marker systems differ in their capacity to detect heterozygotes and may, therefore, differ with regard to the amount of information they can provide. This can further complicate linkage analysis in outbreeding species, as marker pairs may provide different amounts of information for estimation of recombination frequencies and determination of linkage phases of the markers in the two parents. A detailed survey of all possible marker pair configurations and their relative degrees of informativeness is comprehensively discussed in Maliepaard *et al.* (1997).

Several complications associated with linkage analysis in full-sib families can be circumvented via the so-called double pseudo-testcross approach (Grattapaglia and Sederoff, 1994). Here, linkage maps are constructed for each parent separately, i.e. a linkage map of one parent is constructed using only marker alleles that segregate in that parent, while a separate map is constructed for the second parent based on marker alleles that only segregate in the second parent. As this strategy effectively uses the same concept as a testcross, in that segregation of only parent is tested at a time, although both parents are in fact heterozygotes, it is known as a pseudo-testcross. When applied to both parents simultaneously it becomes a double pseudo-testcross. This approach has successfully been applied to genetic mapping studies in several outbreeding species including poplar (Bradshaw *et al.*, 1994; Cervera *et al.*, 2001), eucalyptus (Grattapaglia and Sederoff, 1994), willow (Tsarouhas *et al.*, 2002), larch (Arcade *et al.*, 2000), spruce (Gossilin *et al.* 2002) and olive (la Rosa *et al.*, 2003).

The construction of separate parental maps by the double pseudo-testcross approach does not include potentially important marker alleles that may segregate in both parents. For this reason, it may be preferable to construct an integrated (or consensus) map for the cross. This is facilitated by the use of molecular markers as allelic bridges (Ritter *et al.*, 1990). Multi-allelic markers are more suited to this role than dominant markers (Ritter and Salamini, 1996), which may provide limited information, e.g. in apple (Hemmat *et al.*, 1994) and *Eucalyptus* (Grattapaglia and Sederoff, 1994), where

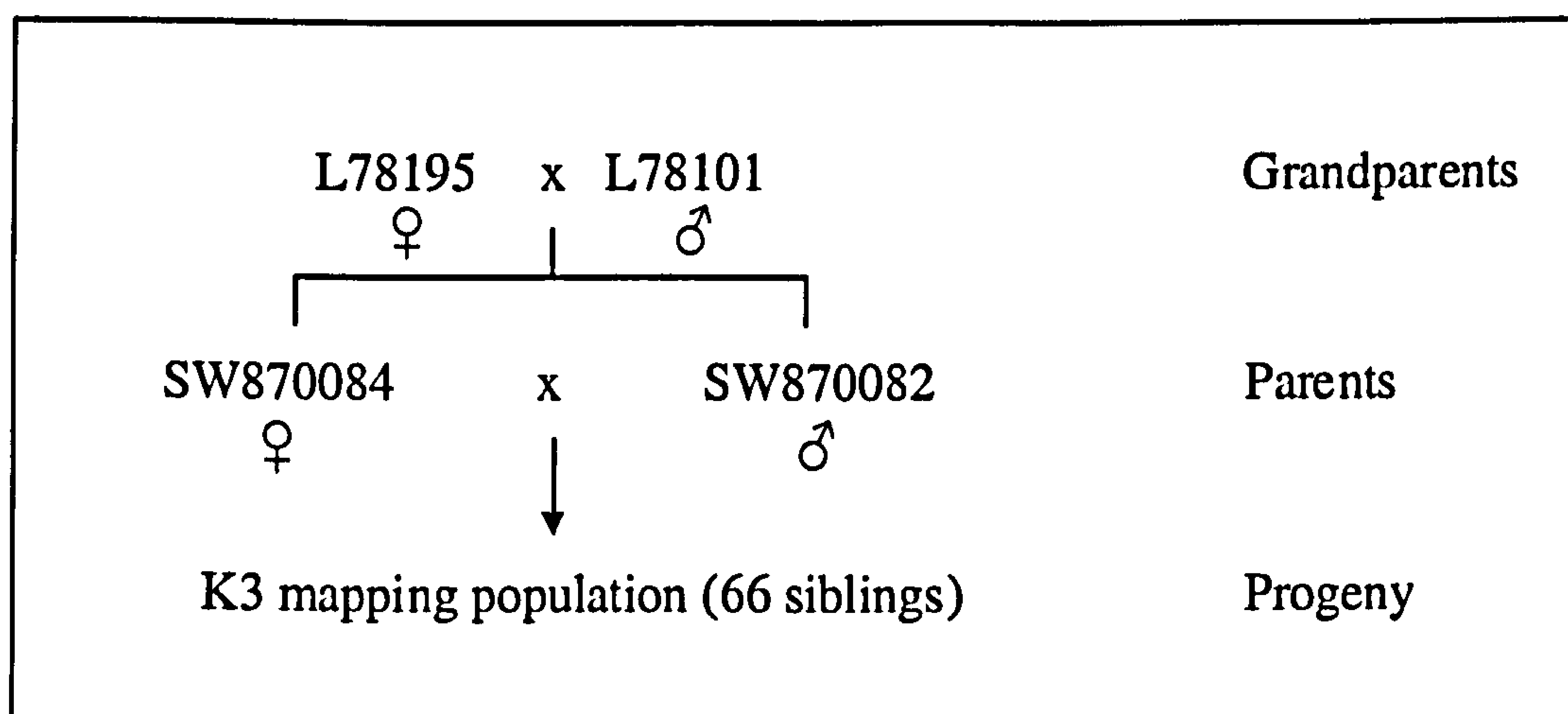
attempts to integrate parental linkage maps based mainly on dominant markers proved problematic. The integration of parental maps produced by the double pseudo-testcross strategy has been reported for several plant species including apple (Maliepaard *et al.*, 1998), rubber (Lespinasse *et al.*, 2000), grapevine (Doligez *et al.*, 2002), maritime pine (Chagne *et al.*, 2002), and now willow (Hanley *et al.*, 2002).

3.2 Materials and Methods

3.2.1 The K3 mapping population

The linkage analysis described in this Chapter is based on the segregating population K3. Originally produced in Sweden by the plant breeding company Svalöf Weibull AB, this full-sib family is the result of a cross between two full-sib parents (SW870084 and SW870082) which are derived from a cross between two non-related, non-inbred, diploid *S. viminalis* clones (L78195 and L78101) (Figure 3.1).

Figure 3.1. Pedigree information for the K3 mapping population



This population was chosen for linkage analysis during the original EU-AIR project as, despite its relatively small size, it was the most suitable population available at that time (J.H.A. Barker, personal communication) and segregated for resistance to *Melampsora* rust (Åhman, 1997).

The mapping population was supplied by Svalöf Weibull AB in the form of 15 cm cuttings, which were planted as part of the UK National Willow Collection at Long Ashton Research Station (LARS) in 1995. Fresh, young leaves were collected from the field site immediately prior to DNA extraction.

3.2.2 DNA extraction

Fresh leaf material was frozen in liquid nitrogen and manually ground to a fine powder prior to DNA extraction using the Nucleon Phytopure Plant DNA Extraction Kit as described in Section 2.4.1. DNA quality and quantity was assessed by agarose gel electrophoresis (Sections 2.4.1 and 2.4.2).

3.2.3 Development of additional AFLP markers

Additional AFLP analysis was performed to supplement the EU AIR project AFLP marker data set. Initially, marker production was based on AFLP Method 1 as described in Section 2.5.2.1. Later, AFLPs were generated according to AFLP Method 2 (Section 2.5.2.2). Details of the primer combinations used are shown in Table 3.3.

3.2.4 Scoring of AFLP markers

Care was taken to only score clearly defined polymorphic bands for inclusion in the linkage data set. Polymorphic AFLP markers were scored manually on a binary basis, i.e. 1 = band present, 0 = band absent, and compiled in Microsoft Excel spreadsheet format. Binary data was subsequently converted in to the format required for analysis by JoinMap™ version 3.0 software as described in Section 2.7.1. All autoradiographs were scored independently by two persons.

The AFLP data set inherited from the EU-AIR project was re-scored, as a significant number of potentially problematic AFLP markers that produced very faint signals or were generally difficult to score, had been included.

3.2.5 Development of additional microsatellite markers

All willow-derived microsatellite markers used in the work described in this Chapter were developed from a microsatellite-enriched, small-insert genomic DNA library previously constructed from *S. burjatica*, Germany (Edwards *et al.*, 1996). The production of markers during the EU-AIR project is described in Barker et al. (2003). For development of additional microsatellite markers during this project, a further 231

microsatellite library inserts were characterised. Of these, 153 clones, previously shown to contain microsatellites by colony hybridisation studies (J.H.A. Barker, unpublished results), were sequenced on an automated ABI 377 DNA sequencer, in both forward and reverse directions, by Dr. Wout Boerjan's group at the University of Gent, Belgium. Resulting data was provided in the form of raw, unedited sequence text files. Remaining clones were sequenced in a single direction using the M13 Forward Universal primer and analysed on an ABI 377 DNA sequencer according to Section 2.4.4). Vector and adapter sequence that flanked the willow DNA insert (see Edwards *et al.*, 1996) was manually identified and removed. Edited sequences were then screened for duplications against previously generated sequence data using Autoassembler™ or Sequencher® software packages.

Primer sets were designed to microsatellite flanking regions (Section 2.5.1.3) if: 1) sufficient sequence length was available between the vector/adapter and the microsatellite motif; 2) primers had not already been designed to a homologous insert; 3) the nucleotide composition of sequence flanking the microsatellite was suitable for primer design. Suitable primers were synthesised and desalted by Sigma-Genosys Ltd..

3.2.6 Microsatellite primer testing and screening

Primer sets were tested for their ability to amplify polymorphic PCR products from the genomic DNA of the parents and grandparents of the K3 mapping population, under the experimental conditions described in Section 2.5.1.4.1. Those that successfully amplified easy-to-score products that were polymorphic between the parents were then used to amplify genomic DNA from the 66 progeny of the mapping population, together with the parents and grandparents for direct comparison on the same gels. PCR conditions were as described in Section 2.5.1.4.1, with the exception of two markers, SB522 and SB565 which were mapped using the fluorescent [F]dCTP labelling technique as described in Section 2.5.1.4.2. Due to time constraints a decision was made that for primer sets that failed to yield a PCR product or produced a banding pattern that was difficult to interpret, no attempt to optimise PCR conditions would be made. Instead, these primer sets were discarded.

3.2.7 Scoring microsatellite markers

All autoradiographs were scored independently by two persons and genotype segregation codes were assigned for each locus in accordance with JoinMap® version 3.0 software requirements (Table 2.1). Analysis of electropherograms resulting from fluorescence-based gels was performed using Genotyper® software.

3.2.8 Poplar microsatellite testing

Nineteen poplar (*Populus*) microsatellite primer sets, developed by Professor. T. Bradshaw (University of Washington, USA) and made publicly available via the Poplar Molecular Genetics Co-operative website (http://poplar2.cfr.washington.edu/pmgc/ssr/pmgcssr.html), were tested on willow DNA. Aliquots of these primers were supplied by Dr. Wout Boerjan, University of Gent, Belgium. Details of these primer sequences are shown in Table 3.1.

Table 3.1. Details of the nineteen poplar microsatellite primers tested in willow.

Locus	Forward primer (5' →3')	Reverse Primer (5' →3')	Motif	Expected product size in poplar
PMGC14	TTCAGAATGTGCATGATGG	GTGATGATCTCACCGTTTG	CTT	210
PMGC61	GATCCCTCTGCACCGTTTAC	ACCCTAAATTTGCTGACAAC	CTT	360
PMGC93	ATCATGCGTTCGGCTACAGC	CTCAAACTCCAACTGTTATAAC	CTT	350
PMGC108	TGCAGGTGATGTCATCACCG	AACCGAATCCATGCGTCACC	CTT	330
PMGC204	GAAGATAAATTCTCCAGCTC	TAACTTTCCCCGCATGT	CTT	225
PMGC223	CGATGAGGTTGAAGAAGTCG	ATATATGTACCGGCACGCCAC	CTT	170
PMGC325	CGATTTATGACAGACAGCTTG	GTACCGTTGAGGTGGCTAG	CTT	295
PMGC333	CTTAGTGGTGAAGTATTC	GAGTGGGTGCTGATTCATCC	CTT	110
PMGC409	ACGTATATGAAGTTCTTGATTGC	GACAGATCATTATGATTACTACAG	GA	150
PMGC451	GCAGCATTGTAGAATAATAAAAG	AAGGGGTCTATTATCCACG	GA	215
PMGC456	AATTACAACCACTTTAGCATATTC	TGCCGACACATCACACATACC	GA	210
PMGC486	TGTAGGAGATATCCACGTGG	AACAATATGCTTCATAGCACAG	GA	115
PMGC510	AGAAGTTGTTGAACCCGATGGG	GCTACAAACTTTGTTGTACCC	GA	150
PMGC576	TAAATTCATGTAGATTGACG	CTTACTATTTCATGGTTGTC	GA	145
PMGC667	CATCCATGATATCAAACCAAATTAG	TGTAATCCAAACATAAAATCCCAAG	GA	115
PMGC684	CCAGCAATGATTGATTGCTCC	GAGCTTTAACTGTCCAGTAGC	GA	260
PMGC2020	TTTTGGCATTCAAAGACTTGGC	AGTTGATTCCATGTCGTGTCC	GA	160
PMGC2021	TAAGGCTCTGTTTGTTAGTCAG	GAGATCTAATAAAGAAGGTCTTC	GA	150
PMGC2098	TCACAAAAGGTTAACGACTTCG	CAGTACTCAGCTGCAGGTCC	GA	180

All primer sets were tested in PCR with genomic DNA from two willow clones (*S. burjatica,* Germany and *S. viminalis,* Astrid) and one poplar clone (hybrid *P. deltoides* x *P. nigra,* Ghoy). PCR conditions were the same as those routinely used for willow (Section 2.5.1.4.1). PCR products were separated by agarose gel electrophoresis using 2% agarose (Section 2.4.2).

3.2.9 Sequencing of poplar microsatellite products at locus PMGC223

This microsatellite locus was selected for direct sequencing as PCR products appeared homozygous (as a single band) in all three samples when visualised by agarose gel electrophoresis. Gel purification and direct sequencing of amplification products from primer set PMGC223 was performed as described in Hanley *et al.* (2000). Resulting sequences were aligned using Autoassembler™ software (Section 2.4.5).

3.2.10 ESTP markers

To test the efficacy of using EST data generated in *Populus* for the production of ESTP markers in *Salix*, two hybrid *Populus*-derived (*P tremula x P. tremuloides*) EST sequences (Sterky *et* al., 1998) were chosen from the public PopulusDB database (http://www.biochem.kth.se/PopulusDB). The first, B006P35U (Genbank accession number: AI166034), was identified as homologous to several phenylalanine ammonia-lyase (PAL) homologous gene sequences, while the second, Q026P52U (Genbank accession number: AI162904), was homologous to known putative receptor-like protein kinase gene sequences. Details of these EST sequences and primers designed for testing in willow are shown in Table 3.2.

Following agarose gel electrophoresis on a 2% gel, amplification products were excised and purified according to Hanley *et al.* (2000) prior to direct sequencing (Section 2.4.3 and Section 2.4.4). For mapping of EST A026P52U (referred to in this thesis as locus ESTP1 from this point forward), genomic DNA of the mapping population was amplified as before and examined by agarose gel electrophoresis. Genotypes were assigned JoinMap® segregation codes (Table 2.1) before inclusion in the willow mapping data set.

Forward and reverse primers were designed for amplification of both EST sequences and used in PCR on genomic DNA of the K3 parents according to the method described for amplification of willow microsatellite loci, with the exception that an annealing temperature of 58°C was used in thermocycling.

Table 3.2. Details of *Populus* EST sequences used in ESTP development. Original EST names (Sterky *et al.*, 1998) are shown above GenBank accession numbers (*in italics*). Regions of EST sequence to which primers were designed are *shaded*.

EST	Sequence	Primers (5'→3')
B006P35U *AI166034*	AATGCAGAAACTAAGGCAAGNTCTAGTTGATCATGCCTTGATGAATGGCGAG AAGGAACAGAATTCAAGCACTTCAATTTTCCAAAAGATTGGAGCCTTTGAGG AAGAACTGAAGAACCTTTTG**CCGAAAGAAGTAGAGAGTGCC**AGACTTGAA CTTGAGAACGGCAACCCGGCTATTCCAAATCGGATCATGGAATGCAGGTCAT ACCCCTTGTACAAGTTTGTGAGGGAAGAATTGGGAACCGTTTTACTAACCGG TGAGAAGGGTCGGATCACCCGGGGAGGAGTTTGACAAGGTATTTACAGCTAT ATGTGCAGGGAAGTTGATTGATCCCATGTTGGAGNGTTTGAAGGAATGGGAA TGG**CGCTCCTCTTCCTCTTTGCTA**	F: CCGAAAGAAGTAGAGAGTGCC R: TAGCAAAGAGGAAGAGGAGCG
A026P52U *AI162904*	TCTTTAAGCTGGGTTCGTAATTGAAAATTGCTGTTGATGCTGCAAAAGGATTG GACTATTTGCACNATGCAAGTGATCCACGAATCATACACCGGGATGTGAAGT GTAGCAATATTTTNTTGGACAAGGAGATGAATGCCNAGGTCTGTGCTTTGGC CTTTNTAAGCAAGTTATGCAGGCAGATGCTACCCATGTGACCACTGTTGTCA AGGGCACTGCAGGCTATCTTGATCCCGAATATTATTCTACCCAACAGCTTAC AGAGAAAAGCGACG**TCTACAGCTTCGGTGTTGTTC**TTCTAGAACTCATCTG CGGGCGAGAACCATTGCGTCGCTCTGGAACTCCAGATTCCTTCAATTTGGTGT TATGGGCGACNACCTTGCAGGCAGGAGCATTTGAAATAGT**GGATGAAAACG** **TAAAGGGAAC**TTTTGATGTGGAAAGCATGAGAAAGGCAGCTATAGTAGCTG TAAGG	F: TCTACAGCTTCGGTGTTGTTC R: GTTCCCTTTACGTTTTCATCC

3.2.11 Map construction

Data analysis was performed on the complete data set, i.e. both segregation data generated during the EU-AIR mapping study, and that generated since, was included. JoinMap® version 3.0 software (Van Ooijen and Voorrips 2001) was used for all aspects of linkage analysis. The software was used to first test markers for segregation distortion using a chi-square test. AFLP markers with highly significant levels of distortion ($P < 0.005$) were excluded from further analyses. Parental linkage maps were constructed using AFLP and microsatellite markers heterozygous in one parent only (JoinMap segregation types, *abxaa* and *aaxab*). In addition, microsatellite markers of types *abxac* and *abxcd* were used but were first separated so that only alleles from either the male or female parent were used for construction of each respective parental map. All marker data, including that for AFLP markers of type *a0xa0* and microsatellite markers of type *abxab*, were utilised for construction of the

integrated map. For all maps, linkage groups were determined using a minimum LOD threshold of 4.0 and map construction performed using the Kosambi mapping function with the following JoinMap parameter settings unless stated otherwise: Rec = 0.40, LOD = 1.0, Jump = 5. Resulting linkage maps were drawn using MapChart 2.1 software (Voorrips 2001). Robustness of the integrated linkage map was tested by comparison of marker orders with separate parental maps.

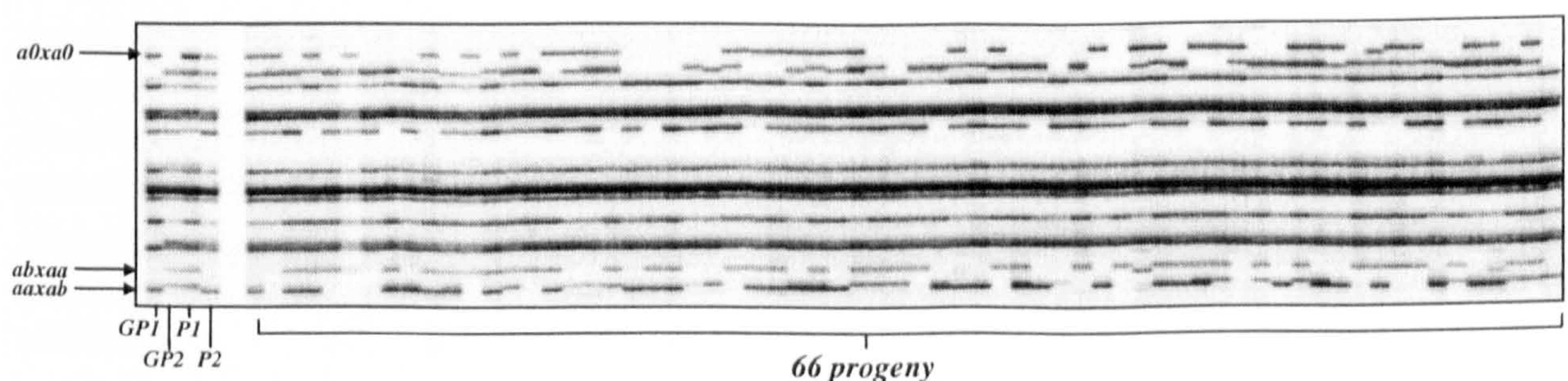
3.3 Results

Results described in this Section concern both the development of additional molecular markers during this project, and subsequent linkage analysis and map construction based on all available marker segregation data.

3.3.1 AFLP markers

In general, both AFLP techniques used were successful in generating high quality marker data for use in linkage mapping. However, the choice of primer combination was shown to greatly influence the quality of the resulting profiles, both in terms of the number of markers detected, and the ease with which they could be scored. For example, primer combination MACA/PCC generated 25 polymorphic markers that could be confidently scored, while combination EACA/MTGT produced only a single useful marker. An example of an AFLP mapping gel generated using the bead selection method is shown in Figure 3.2.

Figure 3.2. A section of an AFLP mapping gel generated with primer combination MACA/PAC. Profiles for the K3 mapping population, along with those of the grandparents and parents are shown. The three possible marker segregation types are also indicated



In total, 477 polymorphic AFLP markers were scored from 48 primer combinations, the characteristics of which are shown in Table 3.3.

Table 3.3. Characteristics of polymorphic AFLP markers generated from 48 primer combinations. Restriction enzymes/adaptors/primers used in AFLP production are indicated for each combination (E = *Eco*RI, M = *Mse*I and P = *Pst*I). Additional characters indicate selective nucleotides. Primer combinations previously screened in the earlier project by Dr J.H.A Barker are shown in *italics*.

Primer combination	Male markers[a]	Female markers[b]	Heterozygous markers[c]	Total polymorphic markers
EAAC/MAAG	4	1	3	8
EAAC/MAAT	0	1	0	1
EAAC/MACC	2	1	2	5
EAAC/MTTG	4	4	0	8
EACA/MACA	5	1	2	8
EACA/MACG	0	0	1	1
EACA/MCTA	2	1	4	7
EACA/MCTC	0	3	1	4
EACA/MTGT	0	1	0	1
EACA/MTTG	5	1	0	6
EACG/MCTA	3	1	2	6
EACT/MAAG	2	2	2	6
EACT/MAAT	0	1	3	4
EACT/MACA	4	3	0	7
EACT/MACC	2	2	0	4
EACT/MGAA	3	6	3	12
EACT/MTTG	0	1	0	1
EAGA/MTGT	1	3	1	5
EAGA/MTTG	2	1	4	7
EAGC/MACC	0	1	4	5
EAGC/MGAA	1	1	4	6
EAGC/MTGT	2	1	3	6
EAGC/MTTG	2	2	2	6
EATA/MTTG	2	2	2	6
MAAG/PCA	9	2	4	15
MAAT/PAC	5	4	10	19
MACA/PAA	7	14	18	39
MACA/PCA	7	6	5	18
MACA/PCC	6	8	11	25
MACC/PAC	4	3	5	12
MACC/PCC	6	4	2	12
MCCG/PAA	2	0	7	9
MCCG/PAC	2	4	0	6
MCCG/PCA	0	2	2	4
MCCG/PCC	1	2	3	6
MCTA/PCA	4	6	6	16
MCTA/PCC	4	3	5	12
MGAA/PAA	4	4	7	15
MGAA/PAC	5	2	6	13
MGAA/PCA	5	4	4	13
MGAA/PCC	4	4	5	13

Table 3.3 continued.

Primer combination	Male markers[a]	Female markers[b]	Heterozygous markers[c]	Total polymorphic markers
MGGC/PAA	2	7	8	17
MGGC/PAC	3	1	1	5
MGGC/PCA	4	2	9	15
MGGC/PCC	5	6	2	13
MTTG/PAC	2	6	6	14
MTTG/PCA	2	4	7	13
MTTG/PCC	4	6	13	23
Total	**143**	**145**	**189**	**477**

a: *abxaa* marker; present in the male parent only; segregating 1:1 in the progeny
b: *aaxab* marker; present in the female parent only; segregating 1:1 in the progeny
c: *a0xa0* marker; heterozygous in both parents; segregating 3:1 in the progeny

The average total number of bands identified per assay was 56 for the bead selection method and 41 for the two-step amplification method. The percentages of scorable polymorphic markers for these two methods were 35% and 18%, respectively. Following analysis of genotype frequencies for each marker, ~15% of markers showed segregation distortion ($P < 0.1$: chi-square test). Distorted markers with P values less than 0.005 (~4% of total markers) were discarded prior to construction of the parental and consensus linkage maps.

3.3.2 Microsatellite markers: development and screening

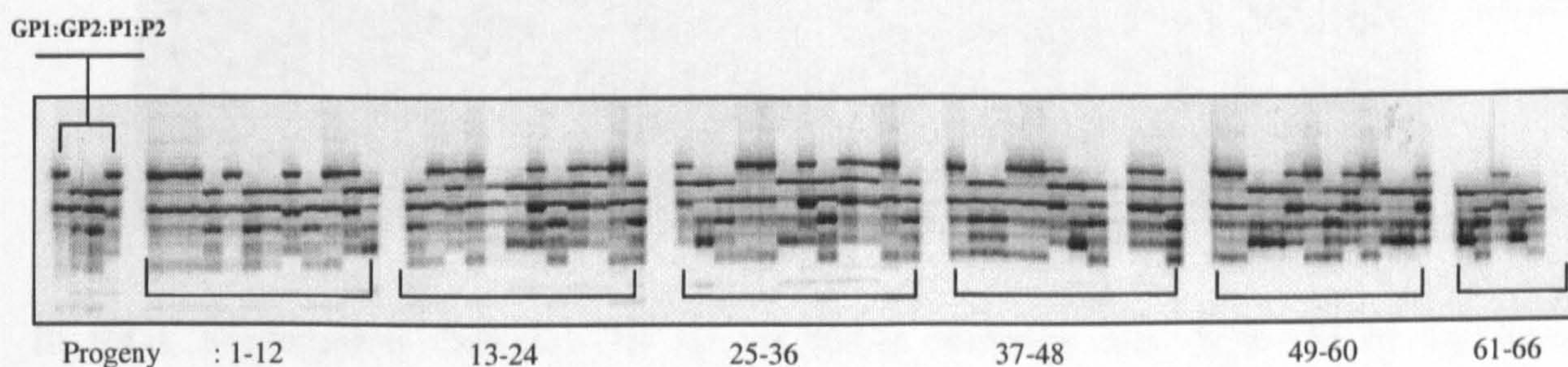
From the 153 microsatellite library inserts sequenced at the University of Gent, 72 primer sets were designed. Primers could not be designed to 44 inserts as the microsatellite motif was too close to the vector/adaptor sequence. A further seven sequences were discarded as were not of sufficient length or the base composition did not permit the design of suitable primers. Finally 27 clones were shown to be duplicates of previously characterised inserts and three contained willow DNA inserts without a microsatellite motif.

An additional 120 clones were sequenced at LARS. Single direction, high quality (i.e. bases could be reliably called in both regions flanking the microsatellite), sequence was obtained for 78 inserts, of which five were duplicates and 13 had microsatellite motifs that were too close to the vector/adaptor sequence to permit primer design. No microsatellite motif was found in eight of the sequenced inserts. Nine inserts were too short to permit primer design. A further 43 primer sets were designed. In total, 115

willow microsatellite primer sets were designed and tested during this phase of the project.

Single locus, easy-to-score, PCR products were observed for 72 (63%) of the loci. Of these, 42 (58%) were found to be polymorphic between the K3 parents and were subsequently mapped. An example is shown in Figure 3.3.

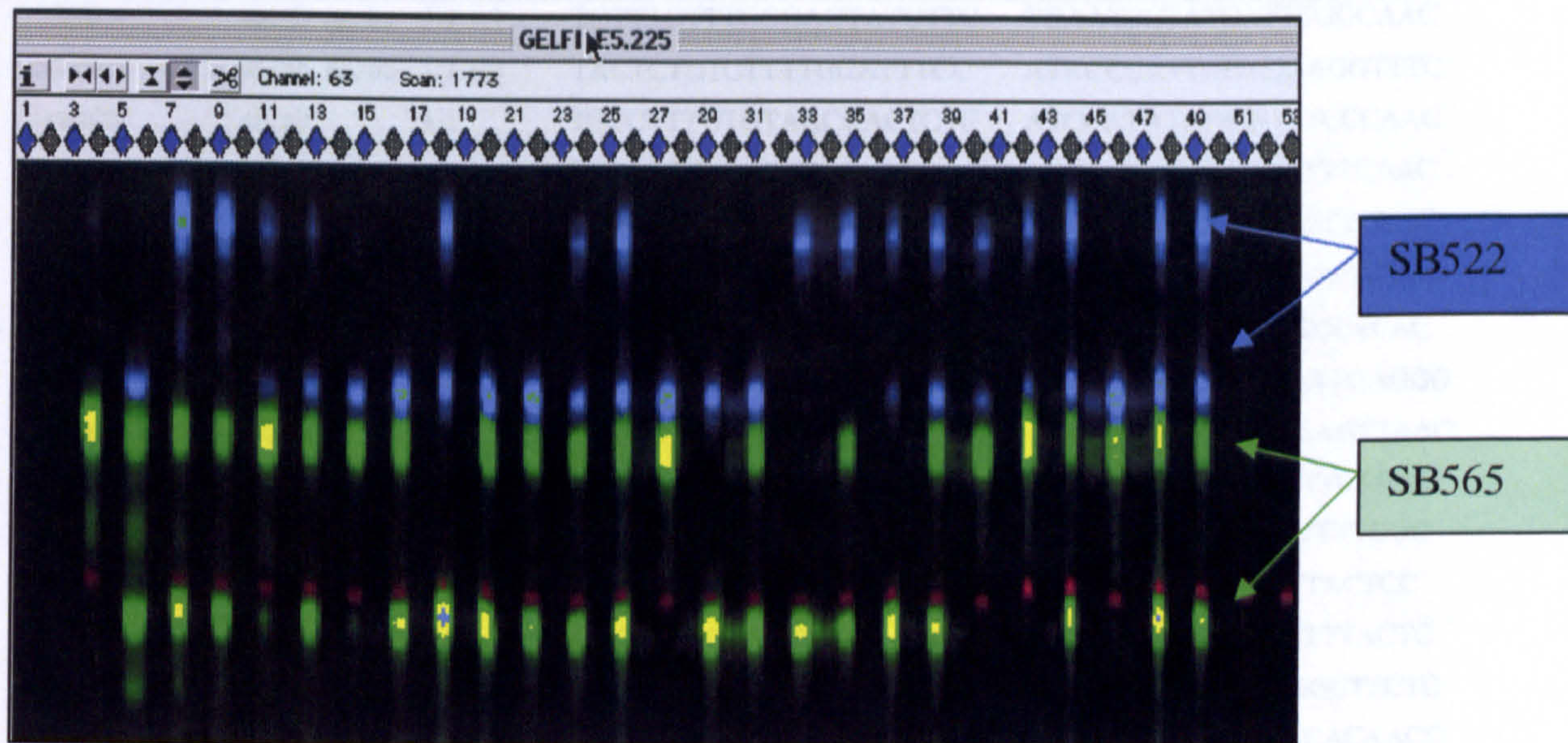
Figure 3.3. Autoradiogram showing segregation of three alleles at polymorphic microsatellite locus SB337.



Twenty primer sets (17%) generated multi-locus products that could not be reliably scored and 23 (20%) primer sets failed to yield a PCR product. Primer set SB442 generated a profile that was difficult to score co-dominantly due to the presence of a greater number of alleles than would be expected in a population of this type, i.e. products were derived from multiple loci. However, one allele, present in both parents, segregated in an approximate 3:1 ratio and was, therefore, scored as a dominant marker (type *a0xa0*) and was included in the final mapping data set. Primer set SB525 failed to generate amplification products in the male parent but was observed as heterozygous in the female. When screened against the mapping population, the segregation profile was consistent with the presence of a homozygous null allele in the male parent suggesting that this locus could be mapped, i.e. segregation type *nnxab*, where n = null allele, a = first maternal allele and b = second maternal allele. Segregation of the larger of the two alleles present in the female parent was scored for inclusion in linkage analysis.

Two microsatellite loci, SB522 and SB565 were successfully mapped using fluorescent techniques based on incorporation of [F]dCTP during PCR. A resulting ABI377 gel image is shown in Figure 3.4.

Figure 3.4. An example of a [F]dCTP mapping gel showing fluoresecent detection of microsatellite alleles at two loci SB522 and SB565, both of segregation type *abxab*.



In total, segregation data for 56 microsatellite markers was included in the final mapping data set, of which, 42 were generated from the current project and 14 were developed and screened by Dr J.H.A Barker during the EU-AIR project (Table 3.4).

Table 3.4. Segregation characteristics of microsatellite markers included in the final K3 mapping data set. Markers developed and screened during the EU-AIR project are shown in *italics*.

Locus	Segregation type	Progeny genotypes	Expected segregation	Forward primer (5' →3')	Reverse primer (5' →3')
SB24	<aaxab>	aa, ab	1:1	ACTTCAATCTCTCTGTATTCT	CTATTTATGGGTTGGTCGATC
SB38	<aaxab>	aa, ab	1:1	CCACTTGAGGAGTGTAAGGAT	CTTAAATGTAAAACTGAATCT
SB53	<aaxab>	aa, ab	1:1	CAGTGACTATCAGAAGATATC	CGGACCCTATTCTTTATTTGG
SB54	<abxab>	aa, ab, bb	1:2:1	CACTGATGAGGGTAATGCTAA	GCACCTTCTGTATAGCCAACC
SB55	<abxaa>	ab, aa	1:1	TAGGGAATCCTAGTTCTTGCA	CTTCAAAGACCCTAACTTGAA
SB80	<aaxab>	aa, ab	1:1	TAATGGAGTTCACAGTCCTCC	ATACAGAGCCCATTTCATCAC
SB85	<abxaa>	ab, aa	1:1	CTCAGCAACTCAATCCAACTA	GTTTGTTAGGGGAGGTAAGAA
SB93	<aaxab>	aa, ab	1:1	GACGCACATACACCATTACAC	TGTTAGAAAATTAGGCACGGA
SB111	<abxab>	aa, ab, bb	1:2:1	GATAGTCAAGGTTGGAGATGG	GGTGGAGAAGAAAAGAGCAGA
SB126	<aaxab>	aa, ab	1:1	TAAACTTGTGTTAGTGAAAGC	TCAGCAACCAACAATCTTCTC
SB194	<abxac>	aa, ab, ac, bc	1:1:1:1	TGTGAGATAAGATTTGTCGGT	CCATAAATAAAAAACGTGAAC
SB210	<aaxab>	aa, ab	1:1	TATAAAGACAAATACCTGGGG	CATCAAAGACTGCTAGAAAGG
SB226	<abxaa>	ab, aa	1:1	TGTTGTGCATAGAGATTTTGT	CTTTTCTCCAATTTTTTTCTG
SB243	<abxaa>	ab, aa	1:1	ATTCCTTTCTTCATCAGTAGC	GACAACGCCATTCACATGACC
SB265	<abxac>	aa, ab, ac, bc	1:1:1:1	ATTAGGGTTTGTTGCTTGGTG	AACATACGTTTCAACGAGAAG
SB268	<abxac>	aa, ab, ac, bc	1:1:1:1	CTAGTTCTGGTGGGGAAGATG	TTCTAAACCATCATTTGGGTG
SB274	<abxac>	aa, ab, ac, bc	1:1:1:1	CCGACCCTCTTTTCTTCTATC	AGCACAGATTGAATGATTGAG
SB276	<abxac>	aa, ab, ac, bc	1:1:1:1	AATAACCACTCCATCTTTCAC	TATTAGTGTTTGCTTTGGAGC
SB287	<aaxbc>	aa, ab	1:1	GTAGGGACTATCAGGGAGCAC	TTAGCTGGGACCTTAAATACG

Table 3.4 continued.

Locus	Segregation type	Progeny genotypes	Expected segregation	Forward primer (5' →3')	Reverse primer (5' →3')
SB331	<abxac>	aa, ab, ac, bc	1:1:1:1	TGTTTGTTTGGGAGTATGTTG	TGAATCCTATCATTTGGCAAC
SB337	<abxac>	aa, ab, ac, bc	1:1:1:1	TACTCTGTCTTTTGCATTTCC	ATACCCATTAGAGAAGGTTTC
SB354	<aaxab>	aa, ab	1:1	TGTCTTTGTGTAACGACTCTG	ATCCATATAGCTGATCCCAAC
SB355	<aaxab>	aa, ab	1:1	TTCATTTTCATCTCGAAGCAC	ATTTAGTGCGGATCTTTCAAC
SB367	<abxab>	aa, ab, bb	1:2:1	CCAAGAACTAATTTCCCAGCC	AGTGTATTCATGTGGTTTAGC
SB408	<aaxab>	aa, ab	1:1	TGACAACGCCATTCACATGAC	TACACAAGTAAGCTGATGTTC
SB419b	<aaxab>	aa, ab	1:1	GCATCCACACACAAAACCAAG	ATTTCTACCCCCTCTCCACAC
SB420	<abxaa>	ab, aa	1:1	TGATGAGGTAAAAAGATGGAG	TAAGGAACACATAATTGAGGG
SB430	<abxab>	aa, ab, bb	1:2:1	CACCCTCATAACAAAAATGGC	CAAATCAGAAAAGAAGTTAAC
SB442§	<abxab>	aa, b-	1:3	ATGAAAATGATCAGTGCCTCG	CTGACCTTCACGCATTCACAC
SB452	<abxac>	aa, ab, ac, bc	1:1:1:1	AGGCCCCAGCTTAAGTTATAC	TTATGACGCAGTTTTTTTGGC
SB493	<abxac>	aa, ab, ac, bc	1:1:1:1	TTTCTGGATCAATGGAGCTTG	CATCTTTCCTTCTTTTACTCC
SB494	<abxaa>	ab, aa	1:1	ATCCCTGGTAATAAGAATGTC	TTCAAGCAACTGGTTTTACTC
SB496	<abxab>	aa, ab, bb	1:2:1	CTCGTGCTTCCACAAATCCTC	ACCCAATAAAAATGGCTTCTG
SB504	<aaxab>	aa, ab	1:1	TGTTGATTAAGATGTGACTGG	ATCATGCCACTAAACACAACC
SB522	<abxab>	aa, ab, bb	1:2:1	TATCTGGTCTGAGGTAGTTTG	AAGCAGGTGGGTGATGATTTG
SB525	<aaxab>	aa, ab	1:1	TGAGGCAAGCAAAATTCAGTC	CTAACACAAATCCATTCCACC
SB532	<abxaa>	ab, aa	1:1	TTGCCTTTCTTGTGGTCAAAC	ATCTTCGCTCCTTACTTTCTG
SB537	<abxab>	aa, ab, bb	1:2:1	TATAGAGGATCGCAAAATCAG	TTCACAATAGGAGGGTATCAG
SB565	<abxab>	aa, ab, bb	1:2:1	GAAAATATAATGCCCAGGAAG	ACAGAACACAGCGACATGAAC
SB578	<abxac>	aa, ab, ac, bc	1:1:1:1	GCATAGGACAGCATACAGGTG	CCTTTTAATTTCAATCCCAGC
SB596	<abxab>	aa, ab, bb	1:2:1	AATTACAAGGATGCCTAATGG	CAAGAATATTTCAACTACTCT
SB602	<abxab>	aa, ab, bb	1:2:1	CACCTCATTGTTATTTACTCC	TTGCTGTTGTTGAGAGATTGG
SB617	<abxaa>	ab, aa	1:1	GCCTCTCCATAACCCAAAAAC	TAGAAAATGATCTGACGATGG
SB768	<abxab>	aa, ab, bb	1:2:1	CTAAGAGCTTTGGTCTTTGTT	CCAACTTCTAACATCAGCACC
SB784	<abxaa>	ab, aa	1:1	GCACAGATAAAAAATTGGTTG	ATATGACTAGGAGGATGTGTT
SB865	<abxaa>	ab, aa	1:1	CCAAATTCACCTAATTACAGC	GCAAGAAAGAGATGCTAGTTC
SB868	<abxab>	aa, ab, bb	1:2:1	TCAGGTTTGTTTTTGCTTCTC	GGACCCAGTAACCCATCAAGC
SB874	<abxac>	aa, ab	1:1	TGCAATCTTACATCCTGTCTC	CACAACTCTCTCTACTACCGG
SB880	<abxac>	aa, ab, ac, bc	1:1:1:1	AAAACACCAGAGAACTGCTAC	TACAACTTCATCTTCCTCTCC
SB896	<aaxab>	aa, ab	1:1	CTGATTACAAGATGAAGGTGG	CAAATTGTATGTATGTGCGGT
SB904	<abxab>	aa, ab, bb	1:2:1	CAGGACCCCCGAATTATGAAC	AGTAACCGCTTCTCTGGCAAC
SB907	<abxcd>	ac, ad, bc, bd	1:1:1:1	CTTACACGCACTCAAATACAC	CTAGGTTTCTAAGGTCAGGTG
SB913	<aaxab>	aa, ab	1:2	TGCTTGTGTAATCTCATGCTC	AGTGAAGGCCTCTCTACCTTT
SB921	<abxab>	aa, ab, bb	1:2:1	CAAAGAAAGACAAGAAAGAGC	ACAGGTGATGATGATAAATCC
SB945	<abxac>	aa, ab, ac, bc	1:1:1:1	TACGCCAACAATCTCTCTTAC	GGGCAGTAGAAACTTACAAGG
SB955	<aaxab>	aa, ab	1:1	ACCACTCTCCAAATCCCTTAC	ATATTTTAACAAGCCACGCTC

§: scored as a dominant marker

Segregation distortion ($P < 0.1$) was observed for 11 (19.6%) microsatellite loci (see Figure 3.8 for further details.

3.3.3 Poplar microsatellite primers

Of the 19 poplar microsatellite primer sets tested on willow DNA, five successfully amplified (PMGC93, PMGC108, PMGC223, PMGC456 and PMGC2020). Loci PMGC93 and PMGC2020 were polymorphic between the K3 parents and were mapped.

Despite initially appearing homozygous by agarose gel electrophoresis, direct sequencing of PCR products for locus PMGC223 indicated that this locus was heterozygous in both the poplar hybrid, "Ghoy", and in *S. burjatica*, "Germany", with polymorphisms being attributed to differences in the number of microsatellite repeat units. The region of the sequence following the microsatellite, determined using direct sequencing, could not be conclusively deduced in these two samples. However, *S. viminalis*, "Astrid", produced a single sequence indicating that this locus was homozygous in this cultivar, as originally predicted. Following removal of ambiguous bases from the sequence of the two heterozygous individuals and bases obscured by unincorporated dye terminators, it was possible to align the resulting edited sequences (Figure 3.5). As expected, sequence data revealed the presence of a $[CTT]_n$ microsatellite motif (compound type) in the poplar and willow samples. Sequence homology was generally high although some differences between poplar and willow were detected, i.e. several base substitutions were present and the poplar microsatellite motif was of greater length than found in the willow sequences.

Figure 3.5. Alignment of edited sequence data for microsatellite locus PMGC223 in *S. viminalis* var. Astrid", *S. viminalis*, "Germany" and poplar hybrid, "Ghoy". *Shaded* regions highlight positions at which all three sequences are equivalent. 'x' denotes that sequence could not be determined. '-' denotes base not present. 'N' indicates an ambiguous base.

```
Astrid      GCCGCGAAAGGGACCAGTGATTCTTTGTCTTTTCCAAGAAACCTAGCGTCACTGGTTCTC
Germany     GCCGCGAAAGAGACCAGTGATTCTTTGTCTTTTCCAAGAAACCTAGCGTCACTGGTTCTC
Ghoy        GCCGCAAAAGAGACCAACAATTCTTTGTCTTTTCCAAGAAACCTANTGTTACTGGCTCTT

Astrid      CTTCTTCTTCACTATCGTCACCTTCTTCTT-------------CACTATCGTCACCTTCT
germany     CTTCTTCTTCACTATCGTCACCTTCTTCTT-------------CACTATCGTCACCTTCT
Ghoy        CTTCTTCTTCACTATCGTCACCTTCTTCTTCTTCTTCTTCCTTxxxxxxxxxxxxxxxxx

Astrid      TCATCTTCTTCTTCAACGACTTCTTCAACCTCATCG
Germany     TCATCxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
Ghoy        xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
```

3.3.4 ESTP markers

Both ESTP marker sets amplified products that manifested as discrete bands of approximately the expected size when examined by agarose gel electrophoresis. Products generated using B006P35U primer set appeared homozygous between the K3 parents while those from A026P52U revealed a size polymorphism. At this locus two bands were observed in each parent. In the progeny these alleles segregated according to an approximate 1:2:1 ratio as would be expected for a single locus heterozygous in both parents, thus, this marker could be mapped.

3.3.5 Construction of the parental maps

Upon initial examination of the segregation data, individuals 37 and 47 of the mapping population were found to contain many missing data points and were, therefore, not included in linkage analysis.

For the paternal map, a total of 132 AFLP markers and 25 microsatellite markers that segregated in the mapping population were included in linkage analysis. At a LOD threshold of 4.0, 77.7 % of markers could be assigned to 21 linkage groups each containing a minimum of three markers. In addition, seven duplets were formed and 20 markers remained unlinked. Marker ordering proved problematic for one linkage group obtained at LOD threshold 4.0 (corresponding to Group V of the consensus map) due to the presence of a single AFLP marker (MAAG/PCA/2) which caused spurious linkage with a sub-group of markers that were unlinked at a LOD threshold of 4.5. At LOD 4.0, the calculated map length of this group was suspiciously large (>200 cM). For these reasons, the problematic AFLP marker was removed and the marker groupings were re-examined. For all other groups, map construction proved straightforward, with no markers presenting problems with respect to goodness-of-fit. The resulting 21 linkage groups spanned 777.5 cM and the average interval between markers was 7.9 cM (Figure 3.6).

For the maternal map, 139 AFLP markers and 33 microsatellite markers were included in linkage analysis. Groupings performed with a LOD threshold of 4.0 resulted in the assignment of 83.7% of markers to 25 linkage groups, with 6 duplets and 16 markers remaining unlinked. Marker orders for all groups were determined without difficulty

with the exception of one group (corresponding to Group III of the consensus map), where two markers, SB419b and EAGA/MTGT/1, could not easily be placed. These markers were not linked at a LOD threshold of 4.5 and were, therefore, discarded. The resulting linkage groups spanned 910.2 cM with an average interval between markers of 8 cM (Figure 3.7).

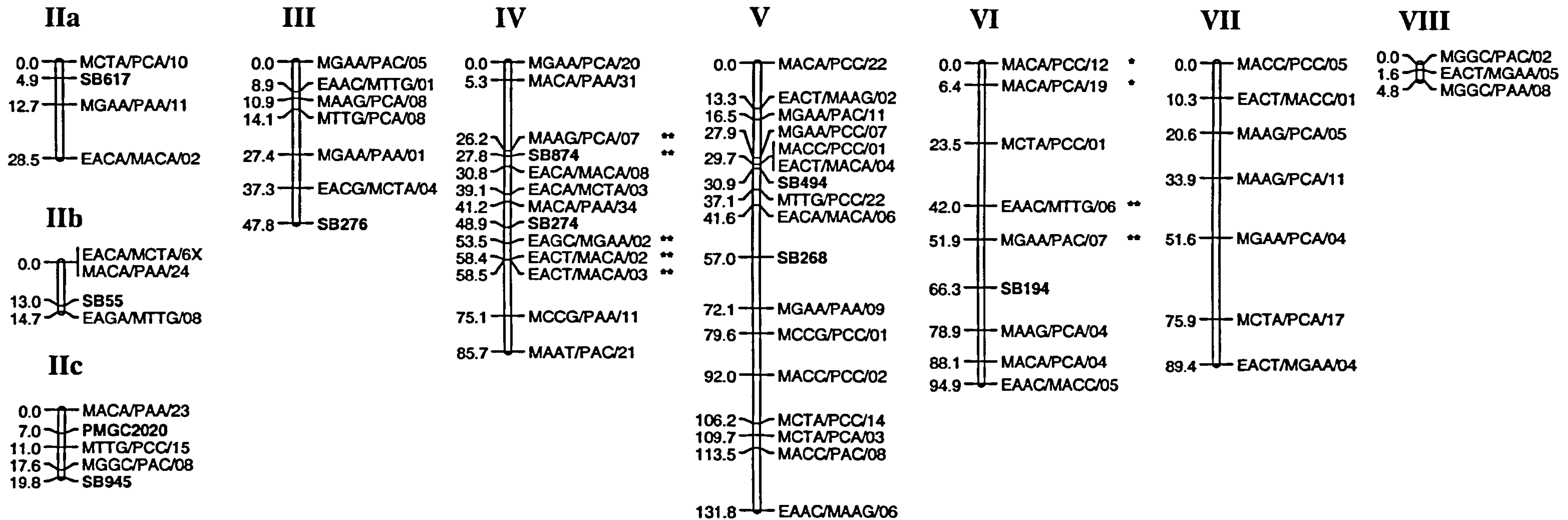


Figure 3.6. The K3 paternal linkage map. Linkage groups containing three or more markers are shown. Linkage group names are assigned according to Hanley *et al.* (2002) where applicable. Marker names are shown on the *right* of each group with map distances (in cM) indicated on the *left*. For details of AFLP marker nomenclature see Table 3.2. Microsatellite markers are indicated in *bold type* with prefixes SB- for those developed from willow and PMGC- from those developed for poplar. Markers showing significant levels of segregation distortion are indicated by *asterisks* (*: $P < 0.1$, **: $P < 0.05$, ***: $P < 0.01$, ****: $P < 0.005$).

Figure 3.6 continued. The K3 paternal linkage map.

†: groups not described in Hanley *et al.* (2002), see text for further explanation

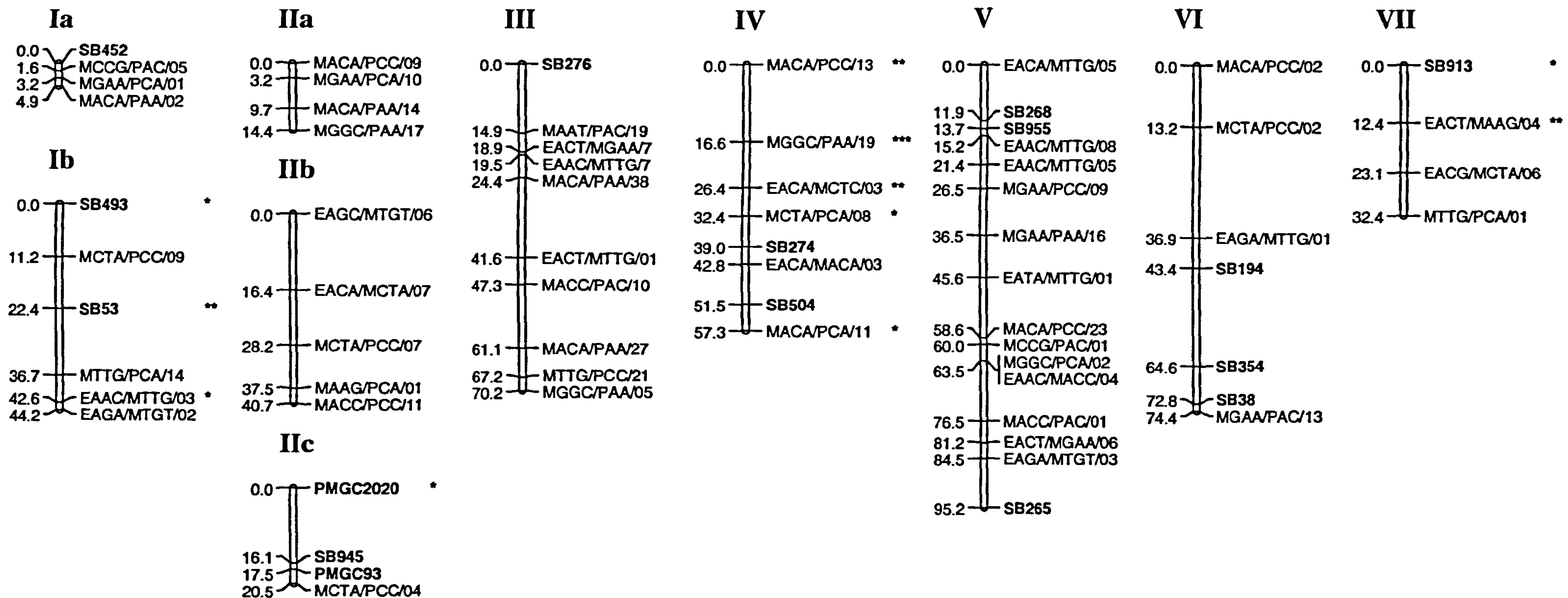


Figure 3.7. The K3 maternal linkage map. Linkage groups containing three or more markers are shown. Linkage group names are assigned according to Hanley *et al.* (2002) where applicable. Marker names are shown on the *right* of each group with map distances (in cM) indicated on the *left*. For details of AFLP marker nomenclature see Table 3.2. Microsatellite markers are indicated in *bold type* with prefixes SB- for those developed from willow and PMGC- for those developed in poplar. Markers showing significant levels of segregation distortion are indicated by *asterisks* (*: $P < 0.1$, **: $P < 0.05$, ***: $P < 0.01$).

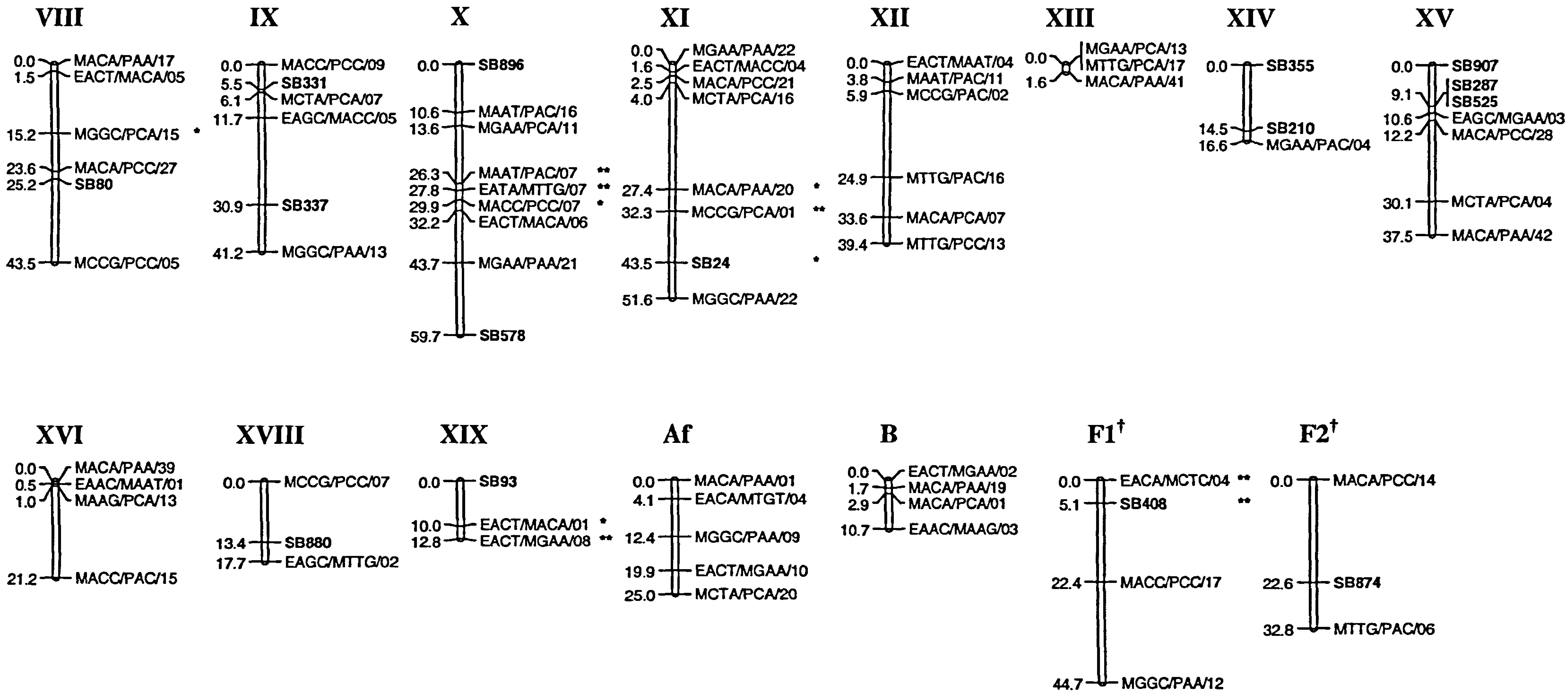


Figure 3.7 continued. The K3 maternal linkage map.

†: groups not described in Hanley *et al.* (2002), see text for further explanation

3.3.6 Construction of the K3 consensus map

The consensus map was constructed from a data set comprising 131 *abxaa* AFLP markers, 139 *aaxab* AFLP markers, 190 *a0xa0* AFLP markers, one ESTP marker and 56 microsatellites of all segregation types. Marker MAAG/PCA/2, which was problematic in construction of the paternal map, was removed prior to linkage analysis.

Initial marker groupings at a LOD threshold of 4.0 resulted in the formation of 36 linkage groups containing three or more markers, with 11 duplets and 40 markers remaining unlinked. However, two groups that were unlinked in the parental maps (X and XIII) were incorporated into a single linkage group in the consensus map. This linkage was deemed to be spurious as these groups showed no evidence of linkage in the parental maps, even at a less stringent LOD threshold of 3.0. Following more detailed examination of linkages for this group, a single *a0xa0* AFLP marker (MGGC/PCA/10) was identified as the cause of this erroneous linkage and was subsequently discarded. When re-analysed, 37 groups containing three or markers were obtained. Two groups (I and II) in the consensus map successfully incorporated previously unlinked groups from the parental maps.

Calculation of recombination frequencies between all marker pairs of the resulting linkage groups led to the identification of a number of 'suspect' linkages in several of the larger linkage groups, i.e. where recombination frequency estimates were greater than 0.6. In all such cases, one marker of each pair was shown to be of the *a0xa0* type. Where highly suspect linkages were observed, as defined by a recombination frequency > 0.7 and a LOD > 1, the *a0xa0* marker involved was discarded from further analyses. For group V, also problematic in linkage group resolution with the paternal map, a suspiciously large number of dubious linkages were highlighted at LOD 4.0. On closer inspection this group was found to contain sub-groups of markers that showed no linkage in the parental maps, e.g. group F2 of the maternal map. An increase in LOD threshold to 7.0 was required to split this group into statistically more robust groups with no suspect linkages. Consequently, the resulting consensus group V does not represent all markers assigned to the corresponding groups in the parental maps.

Marker orders were determined only for those linkage groups that could be identified in one or both parental maps. When corresponding linkage groups of the parental and consensus maps were aligned, the map orders for the majority of markers present in both a parental map and the consensus map were identical for most linkage groups. However, for groups where marker orders were not in complete agreement between maps, the effect of each included AFLP marker of type *a0xa0* was systematically examined. In all such cases, the removal of one or more markers of this type corrected inconsistencies between the parental and consensus maps and reduced 'tension' between markers with regard to goodness-of fit. All remaining markers were easily placed without using the capability of JoinMap software to force problematic marker on to the map. However, for linkage group XI, the position of two AFLP markers remained inconsistent between the maternal and consensus maps. Here, the 'fixed order' feature of JoinMap was used to force the more robust maternal map order on to the consensus group prior to map calculation. Following implementation of this approach, mean chi-square goodness-of-fit contributions for each marker were examined and found to be statistically acceptable.

Microsatellite SB874 could not be placed on the consensus map. This marker detected three alleles in the K3 population (segregation type *abxac*). When scored for parental alleles separately in the parental data sets, this marker could be assigned to linkage group IV of the paternal map. However, this marker showed no evidence of linkage to Group IV of the maternal map, instead forming a separate group with three AFLP markers (group F1). When constructing the consensus map, parental alleles were scored separately and mapped as two independent loci. SB874p1 (paternal) mapped to group IV of the consensus map while SB874p2 (maternal) showed no linkage with any markers other than those identified in maternal group F1.

Linkage groups Am, Af, B, D and E contained only markers derived from a single parent and, hence, no consensus linkage groups could be calculated. In addition, a consensus linkage group for group XVIII, previously described in Hanley *et al.* (2002), could not be calculated. Since this group was originally published, the use of more advanced sequence analysis software has led to the discovery that microsatellite markers SB78 and SB880 both target the same locus. Segregation data for microsatellite SB78 was discarded from the mapping data set, in turn, preventing the

calculation of a consensus group. The corresponding parental linkage groups are now referred to as XVIIIm and XVIIIf for the paternal and maternal maps, respectively.

The resulting 21 K3 consensus groups[‡] comprised 191 AFLP markers of types *aaxab* and *abxaa*, 81 AFLP markers of type *a0xa0*, one ESTP marker and 43 microsatellites of all types (Figure 3.8). Markers showing segregation distortion were detected on eleven of the linkage groups with some evidence of clustering on groups IV, VI, VIII, XI, XIV, XVI and XIX. The total map distance for consensus groups[‡] was 1162 cM with an average marker interval between markers of 4.2 cM.

[‡] defined here as groups previously identified in the parental maps but comprising at least one marker heterozygous in both parents

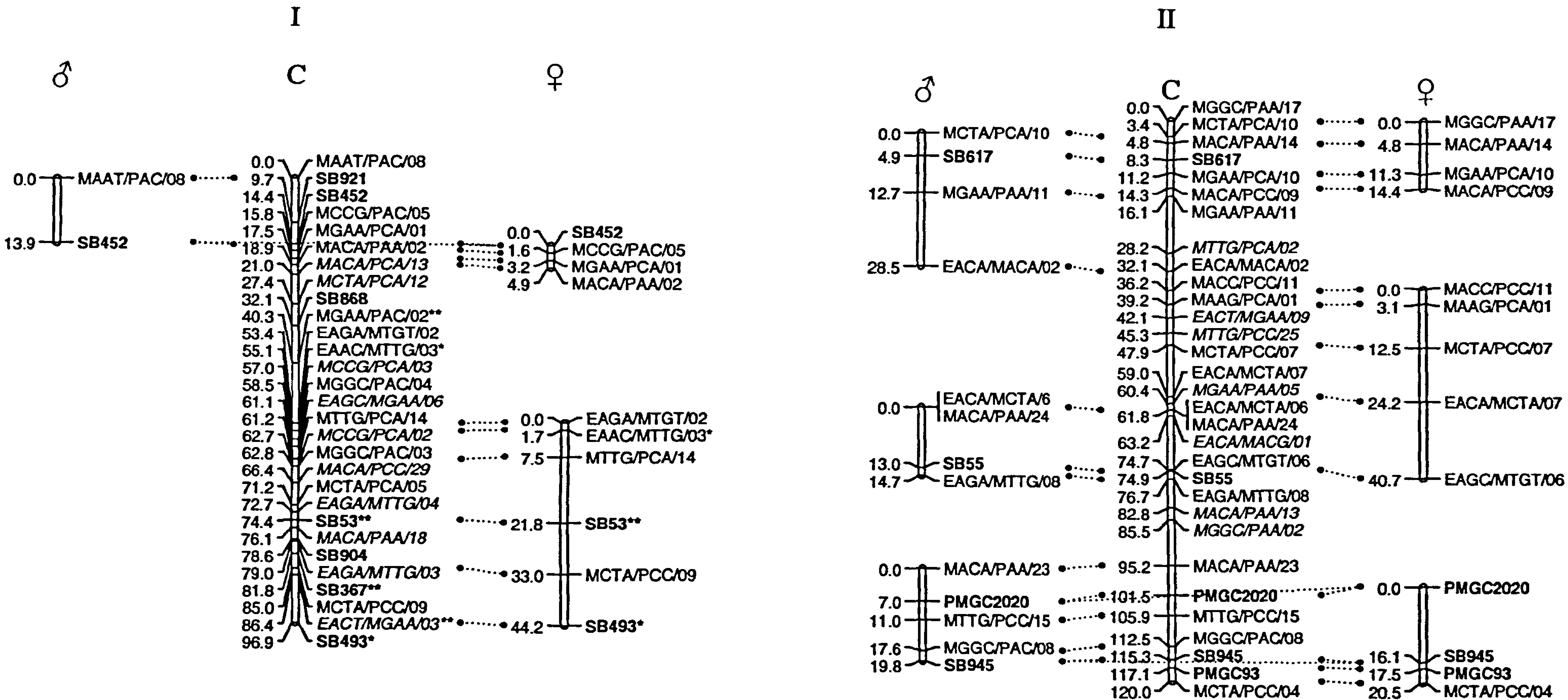


Figure 3.8. The K3 consensus linkage map. Alignments between paternal (♂), maternal (♀) and consensus (C) linkage groups are shown. Linkage group names are assigned according to Hanley *et al.* (2002) where applicable. Marker names are shown on the *right* of each group with map distances (in cM) indicated on the *left*. Microsatellite markers are indicated in *bold type* with prefixes SB- for those developed from willow and PMGC- for those developed in poplar. For details of AFLP marker nomenclature see Table 3.2. AFLP markers of type *a0xa0* are shown in *italics*. Corresponding marker loci are indicated by *dotted bars* between maps. Markers showing significant levels of segregation distortion are indicated by *asterisks* (*: $P < 0.1$, **: $P < 0.05$, ***: $P < 0.01$, ****: $P < 0.005$, *****: $P < 0.001$).

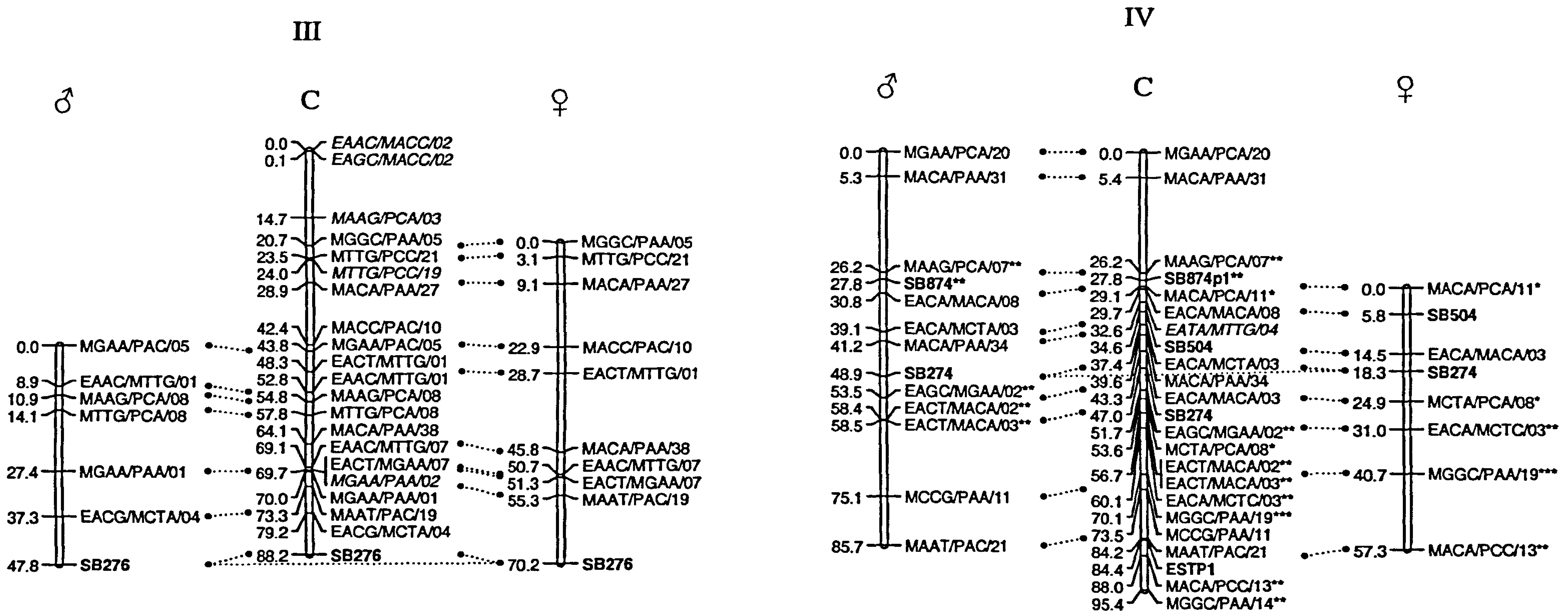
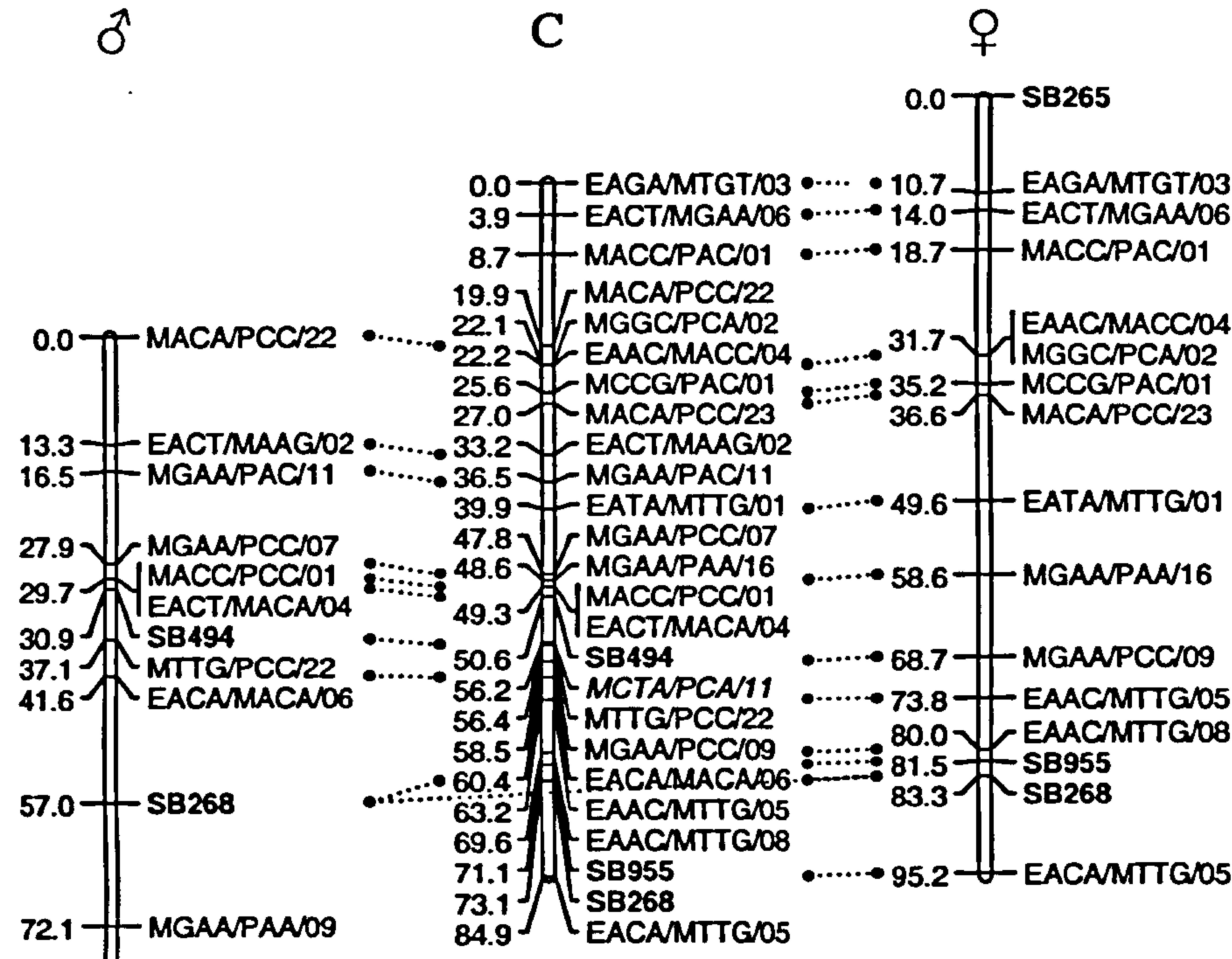


Figure 3.8 continued. The K3 consensus linkage map.



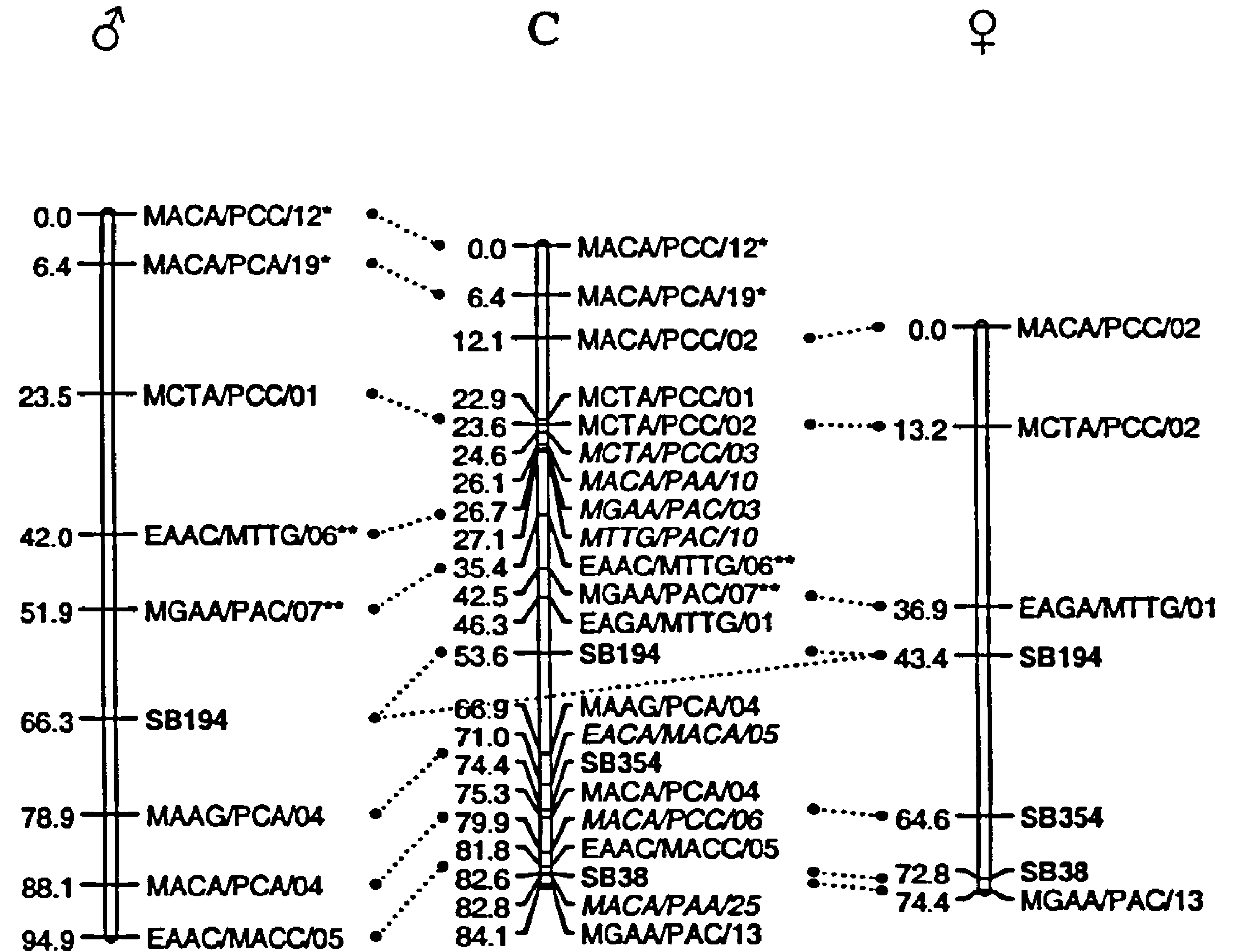
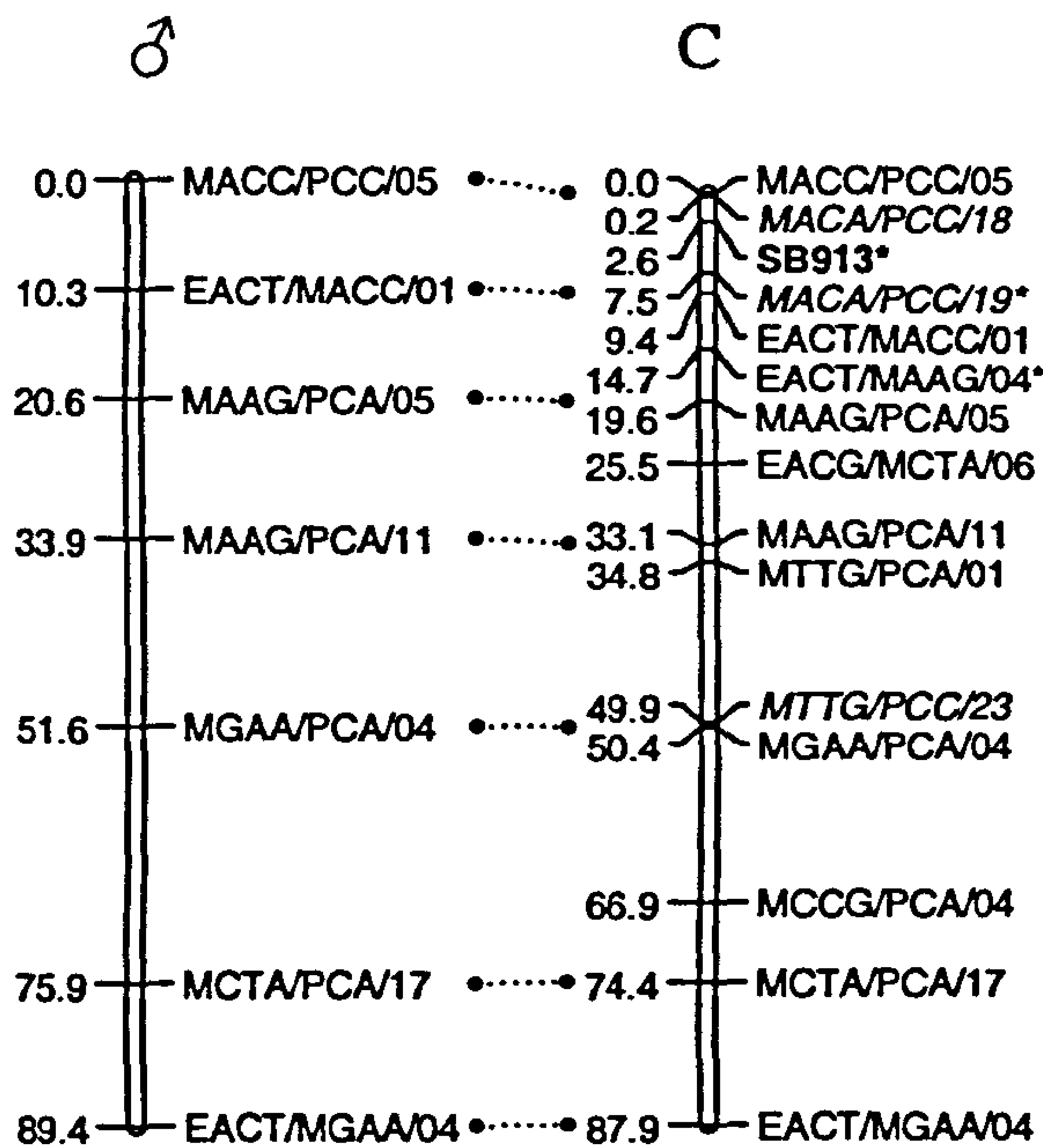


Figure 3.8 continued. The K3 consensus linkage map.



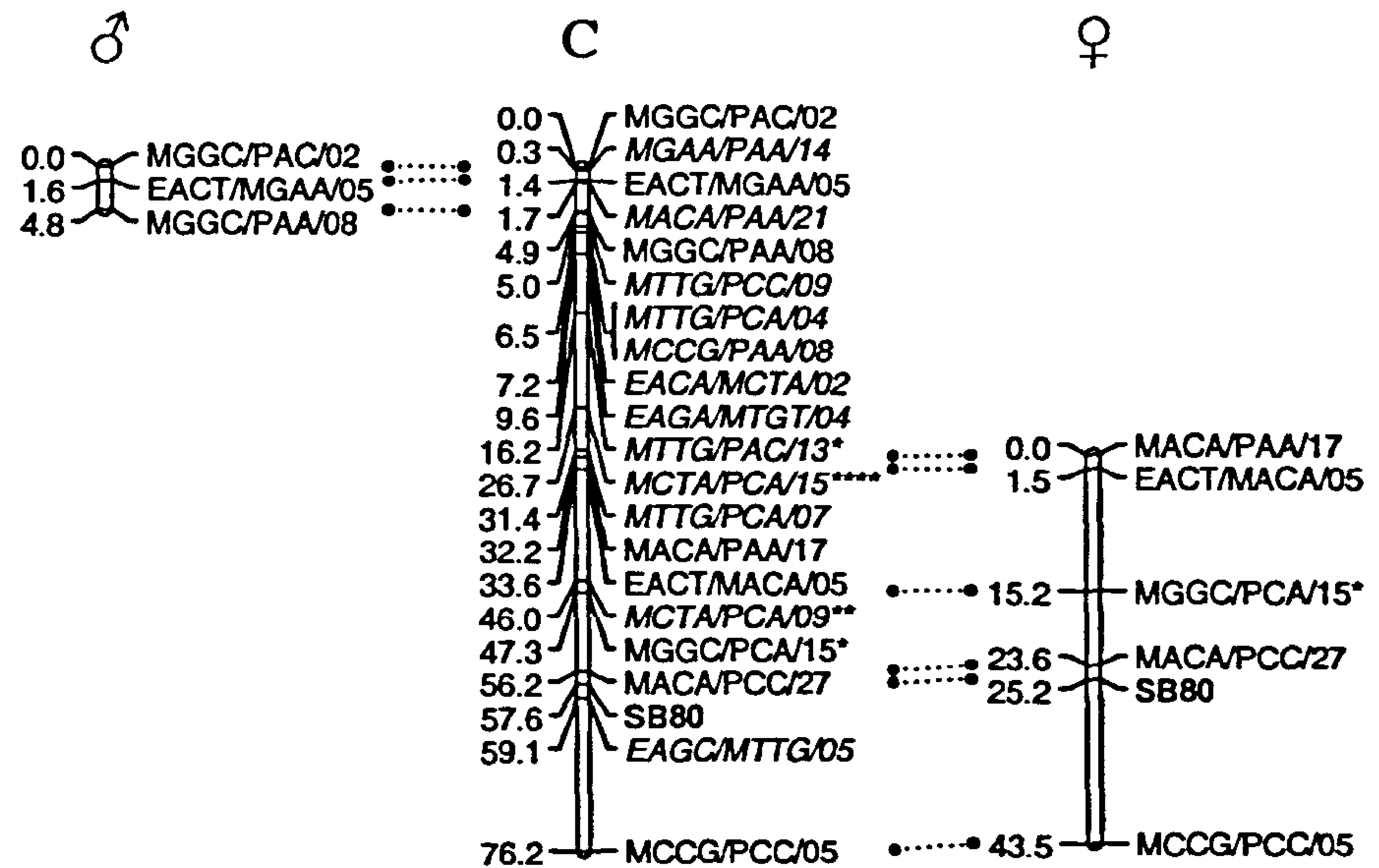
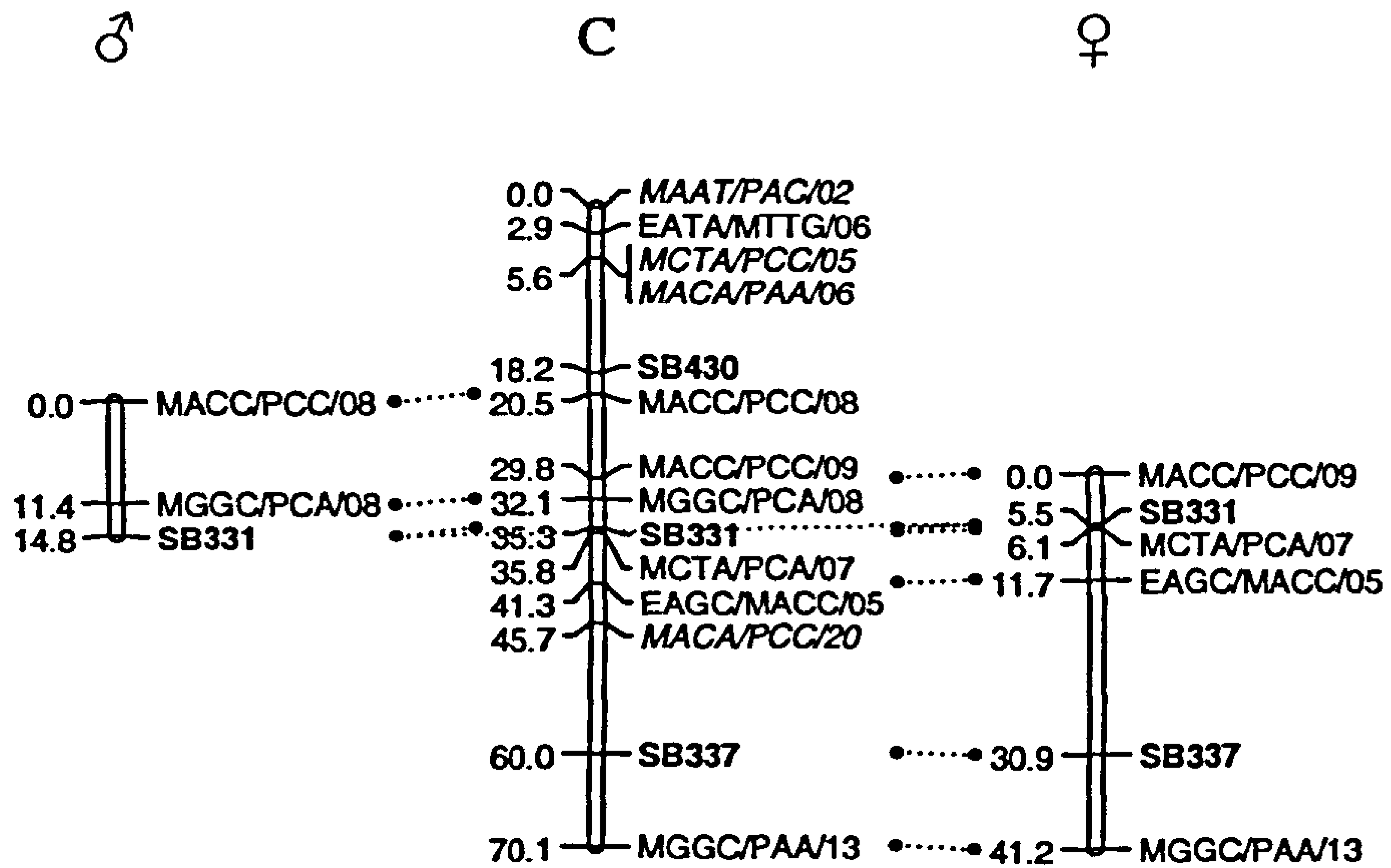


Figure 3.8 continued. The K3 consensus linkage map.



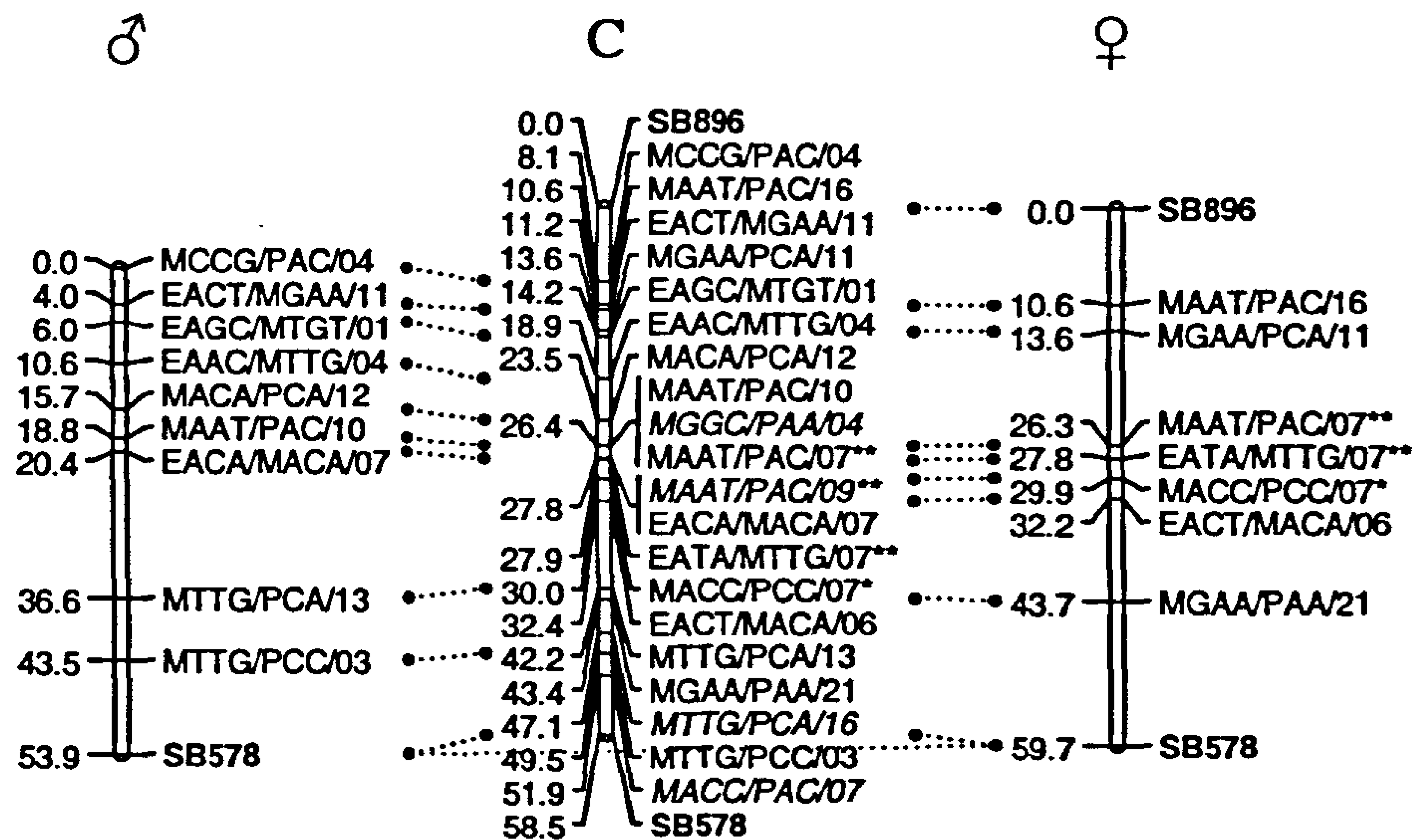
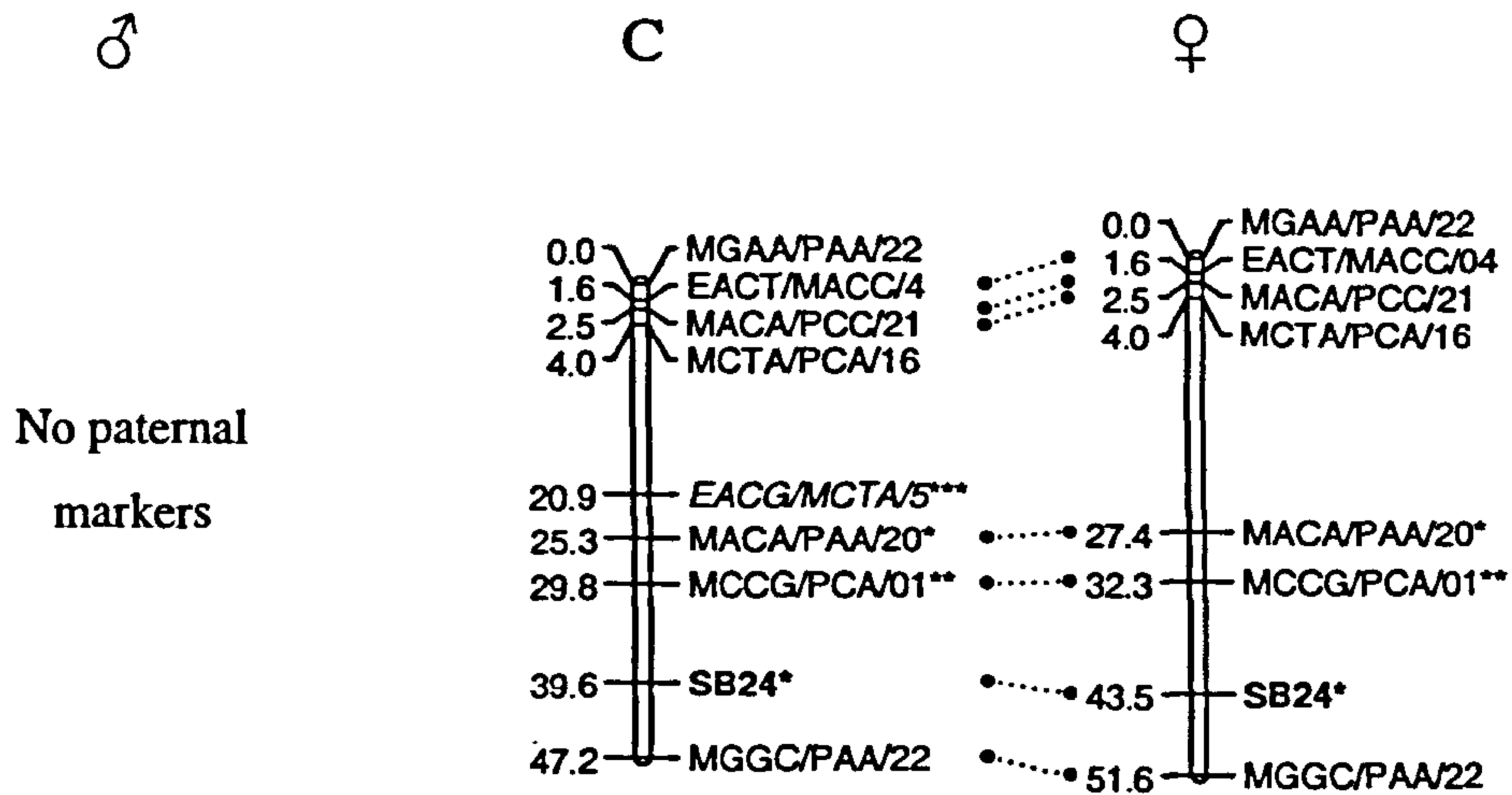
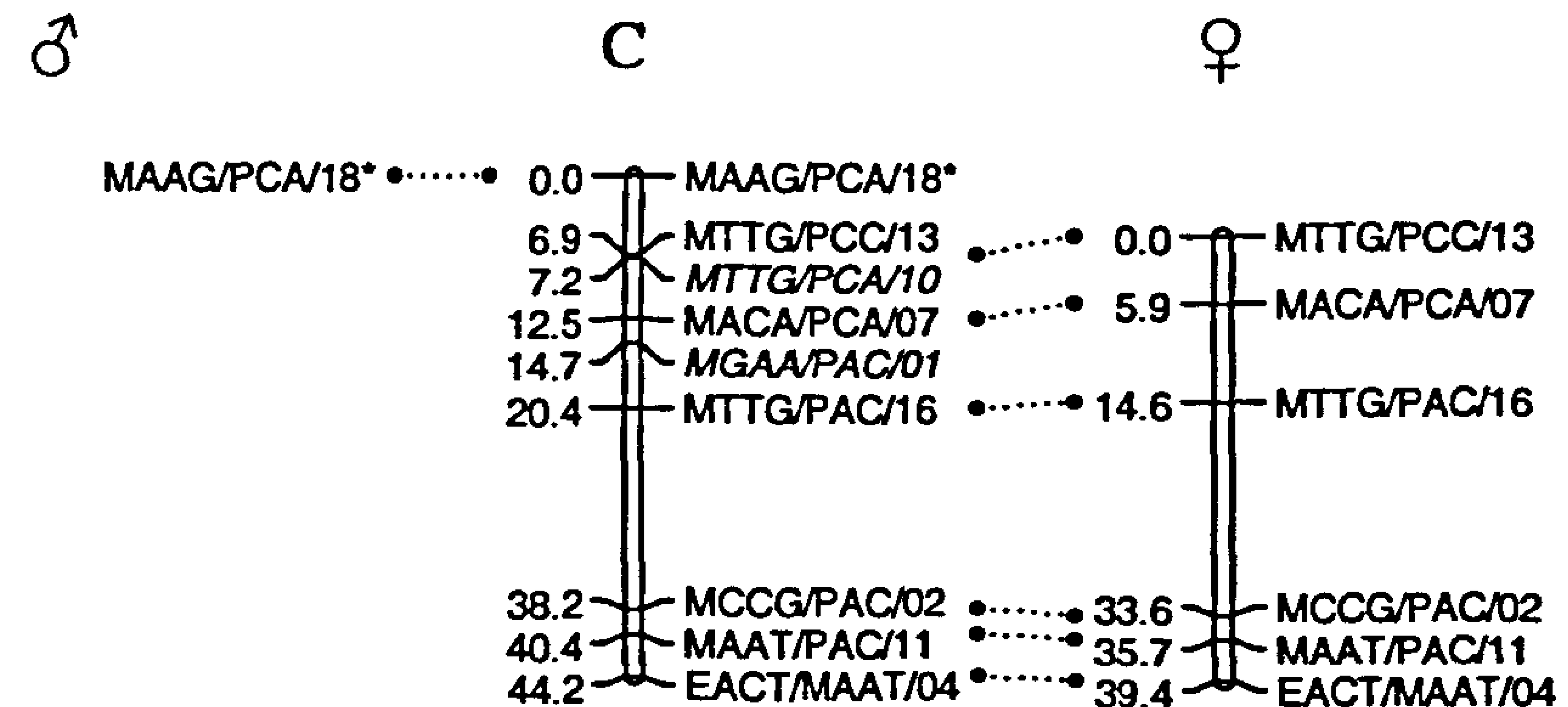
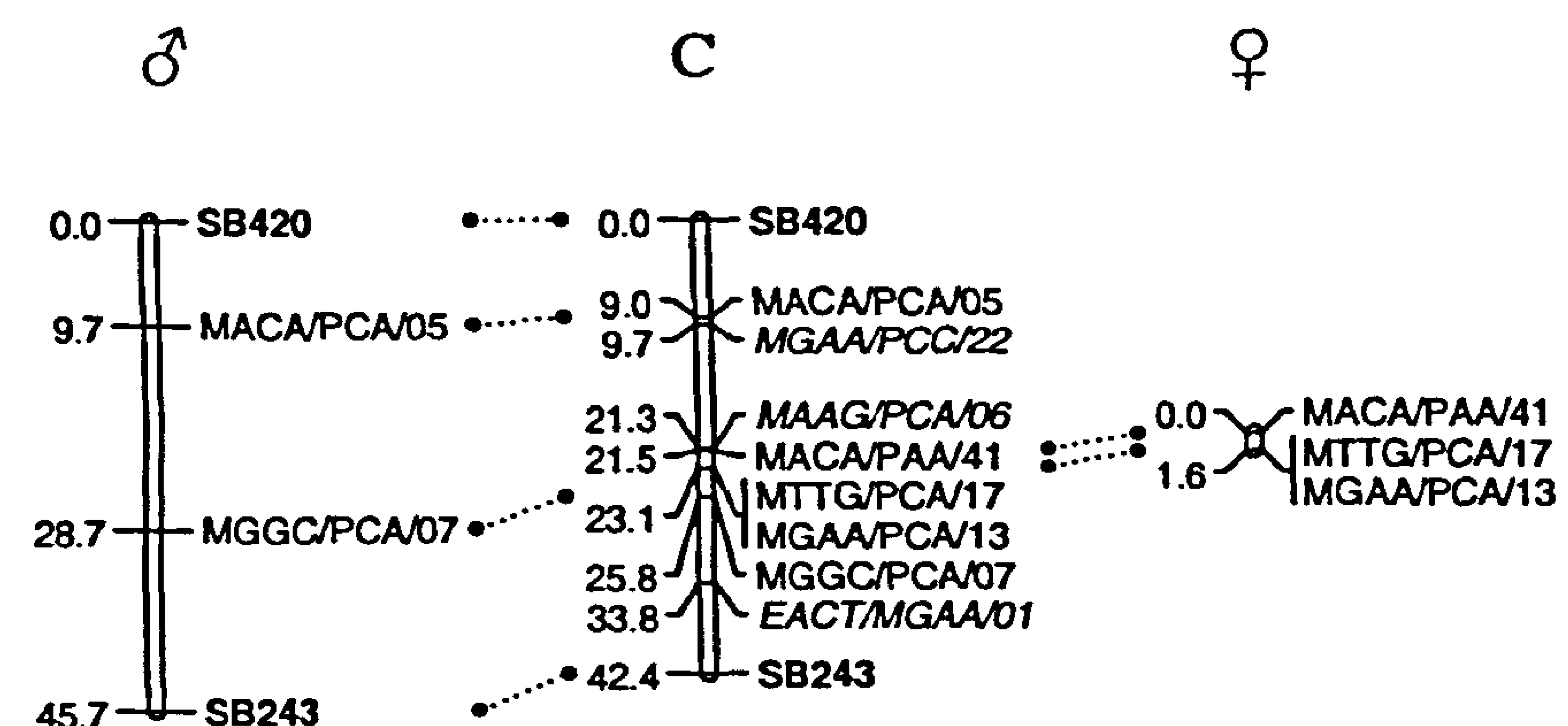


Figure 3.8 continued. The K3 consensus linkage map.







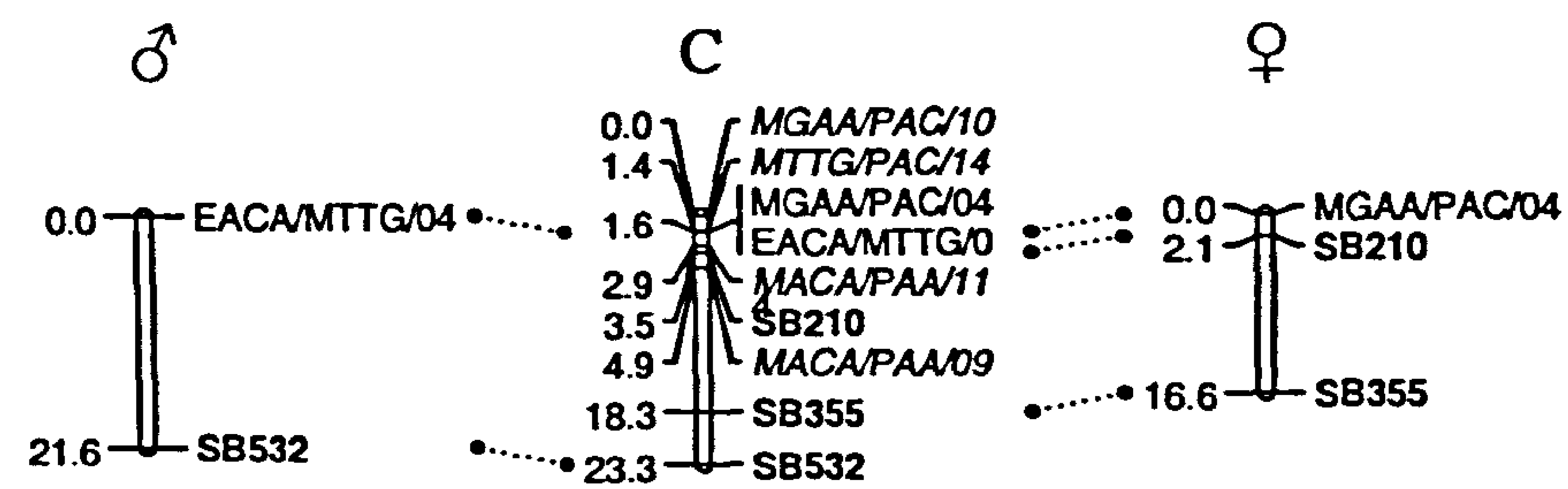
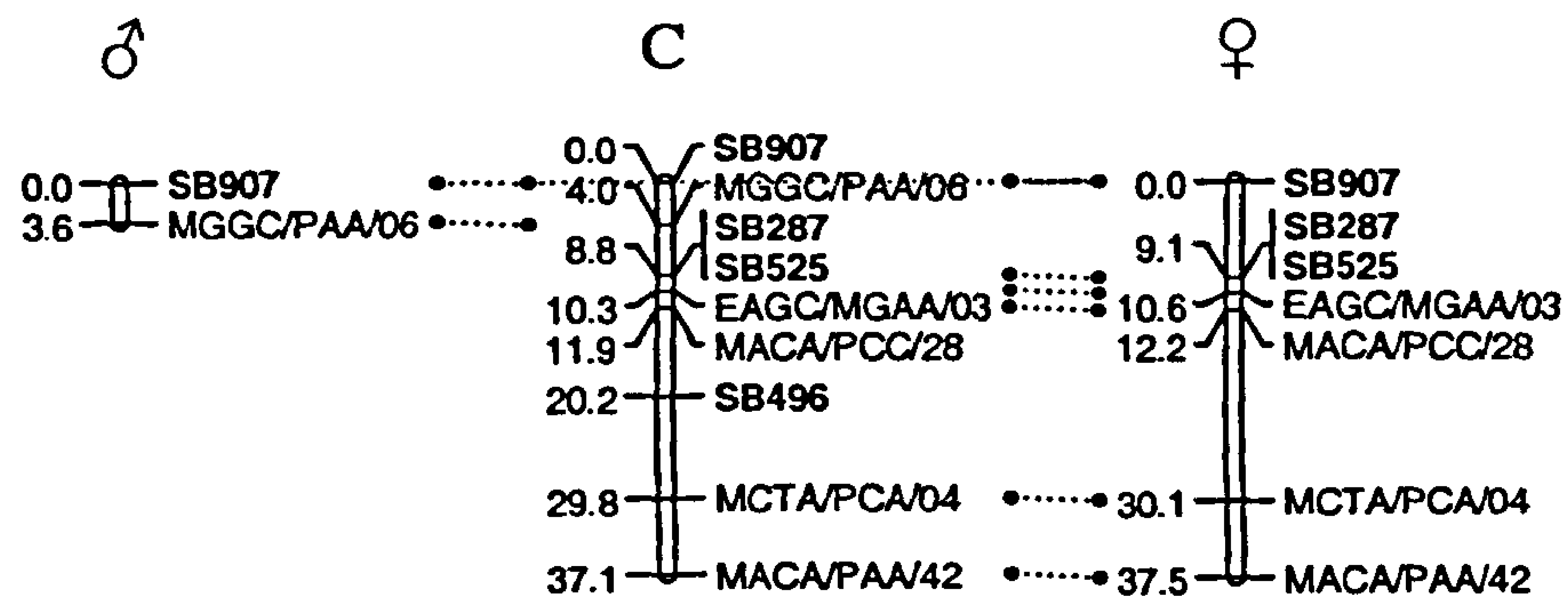
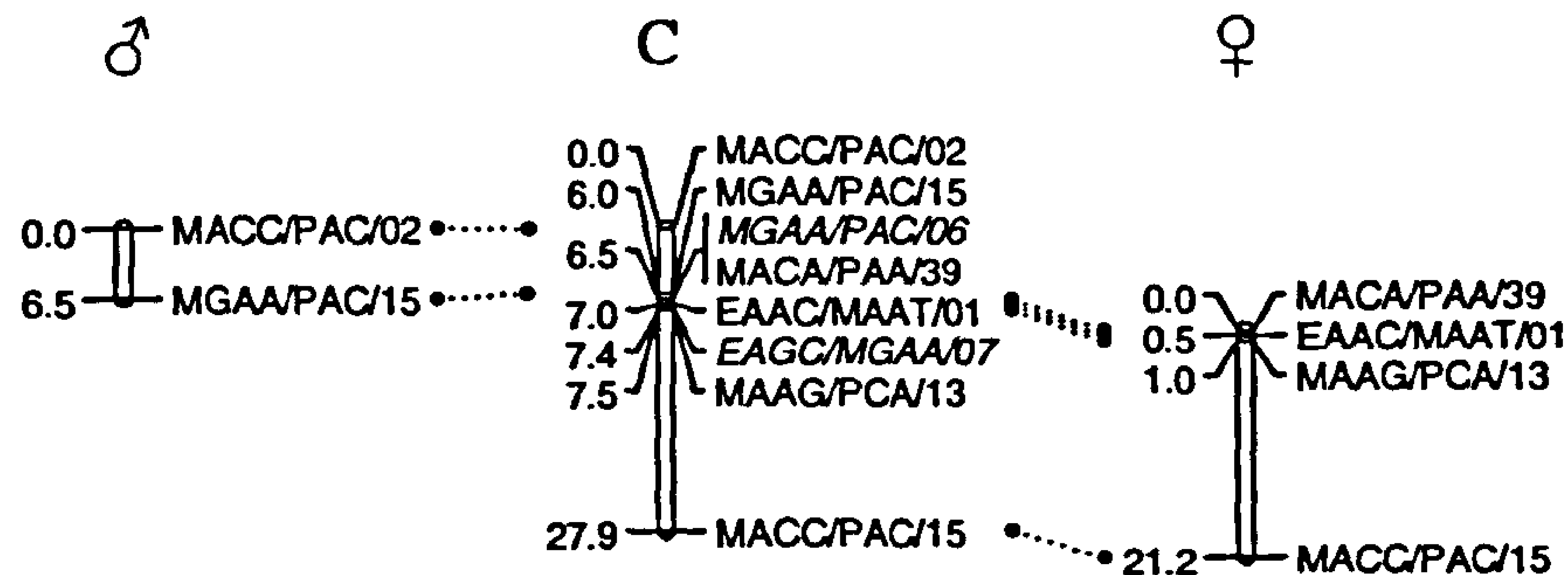
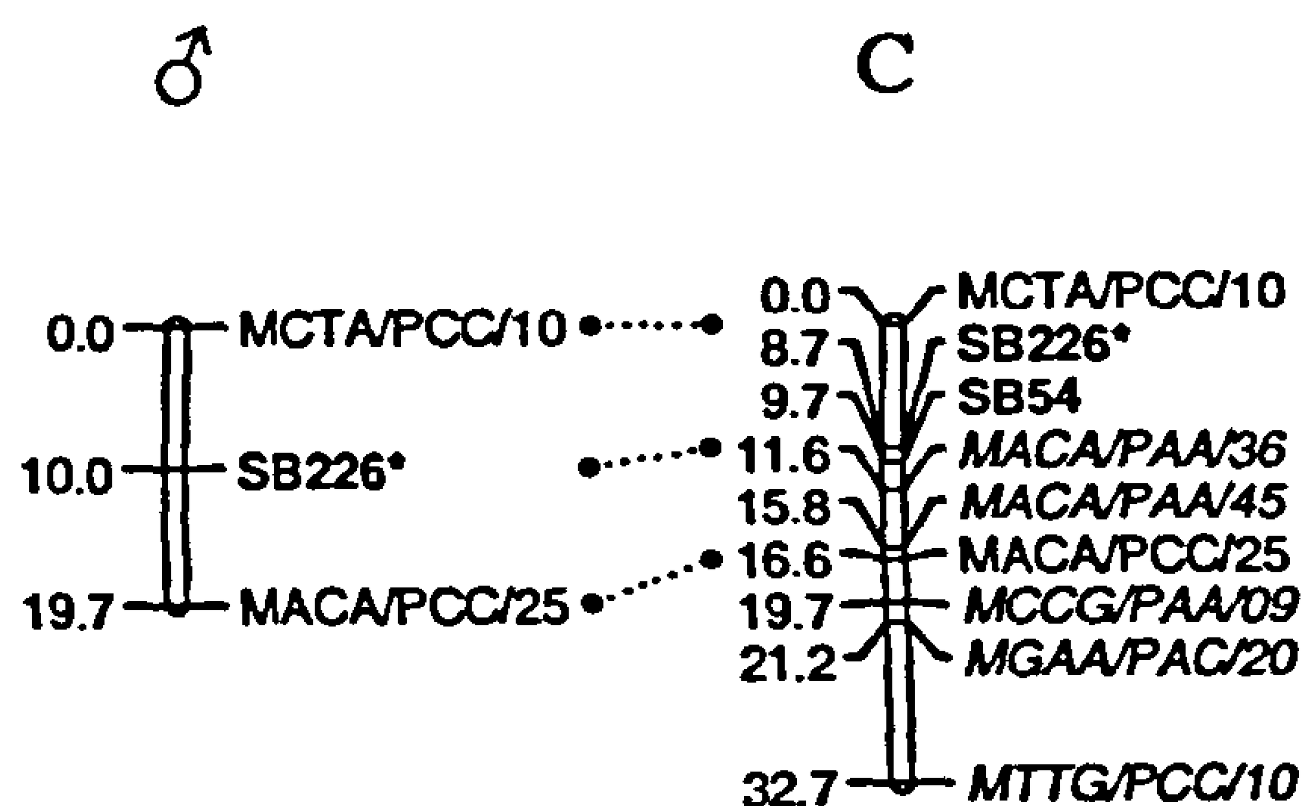


Figure 3.8 continued. The K3 consensus linkage map.







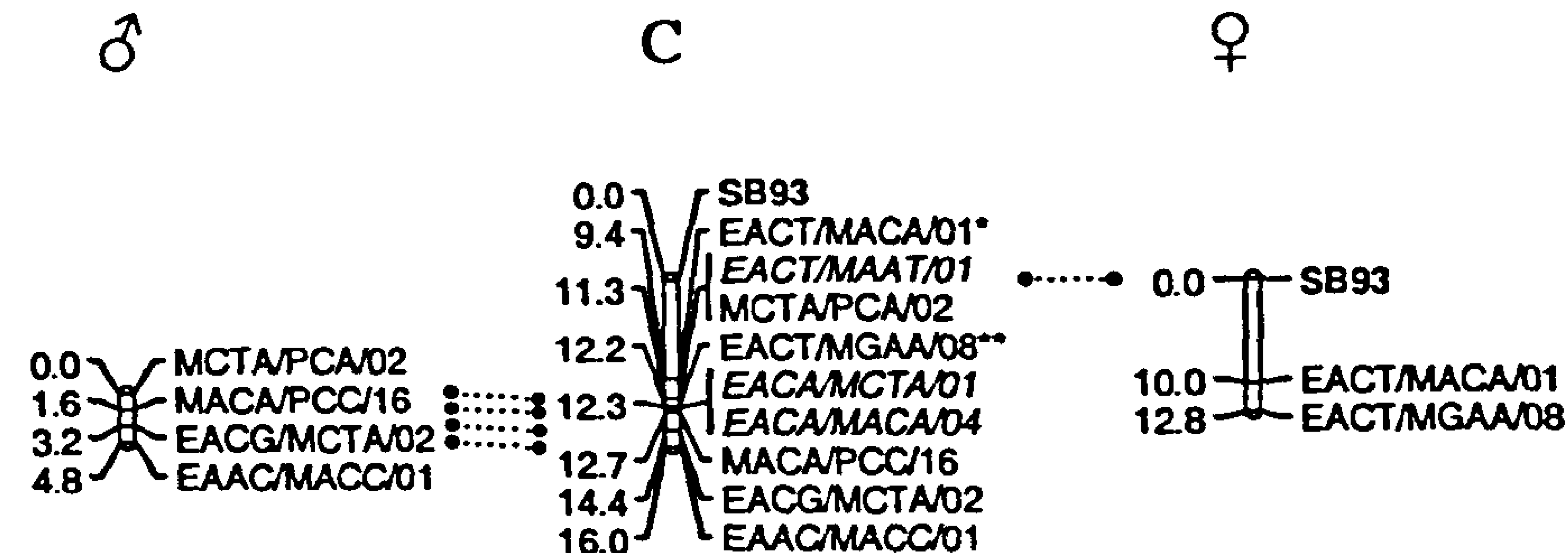


Figure 3.8 continued. The K3 consensus linkage map.

XX

♂

0.0 MACA/PCC/04
2.9 MGGC/PCA/17
6.2 EACT/MAAG/06
7.4 EAGA/MTGT/05
11.4 EAGC/MTTG/01

C

0.0 MACA/PCC/04
3.1 MGGC/PCA/17
3.9 MTTG/PCC/24
5.3 *MGGC/PAA/15*
MGAA/PAA/23
6.2 EACT/MAAG/06
7.4 EAGA/MTGT/05
8.3 SB522
11.1 EAGC/MTTG/01

♀

MTTG/PCC/24

XXI

♂

0.0 EACG/MCTA/07
1.6 MACC/PCC/15
4.7 MGAA/PCC/08*

C

0.0 EACG/MCTA/07
1.5 MACC/PCC/15
MTTG/PAC/04
1.7 MACA/PCA/17
MACC/PAC/11
4.4 MGAA/PCC/08*

♀

MACA/PCA/17

XXII

♂

0.0 MGAA/PCA/14
0.4 MAAG/PCA/15
9.1 MACA/PAA/26
EATA/MTTG/02

C

0.0 MGAA/PCA/14
0.4 MAAG/PCA/15
9.2 MACA/PAA/26
9.3 EATA/MTTG/02
9.8 *EATA/MTTG/03*

♀

No maternal markers

XVIII

♂XVIIIm

0.0 SB880
9.9 EACT/MACA/07
11.5 MACA/PAA/35

C

No consensus group

♀XVIIIf

0.0 MCCG/PCC/07
13.4 SB880
17.7 EAGC/MTTG/02

Figure 3.8 continued. The K3 consensus linkage map.

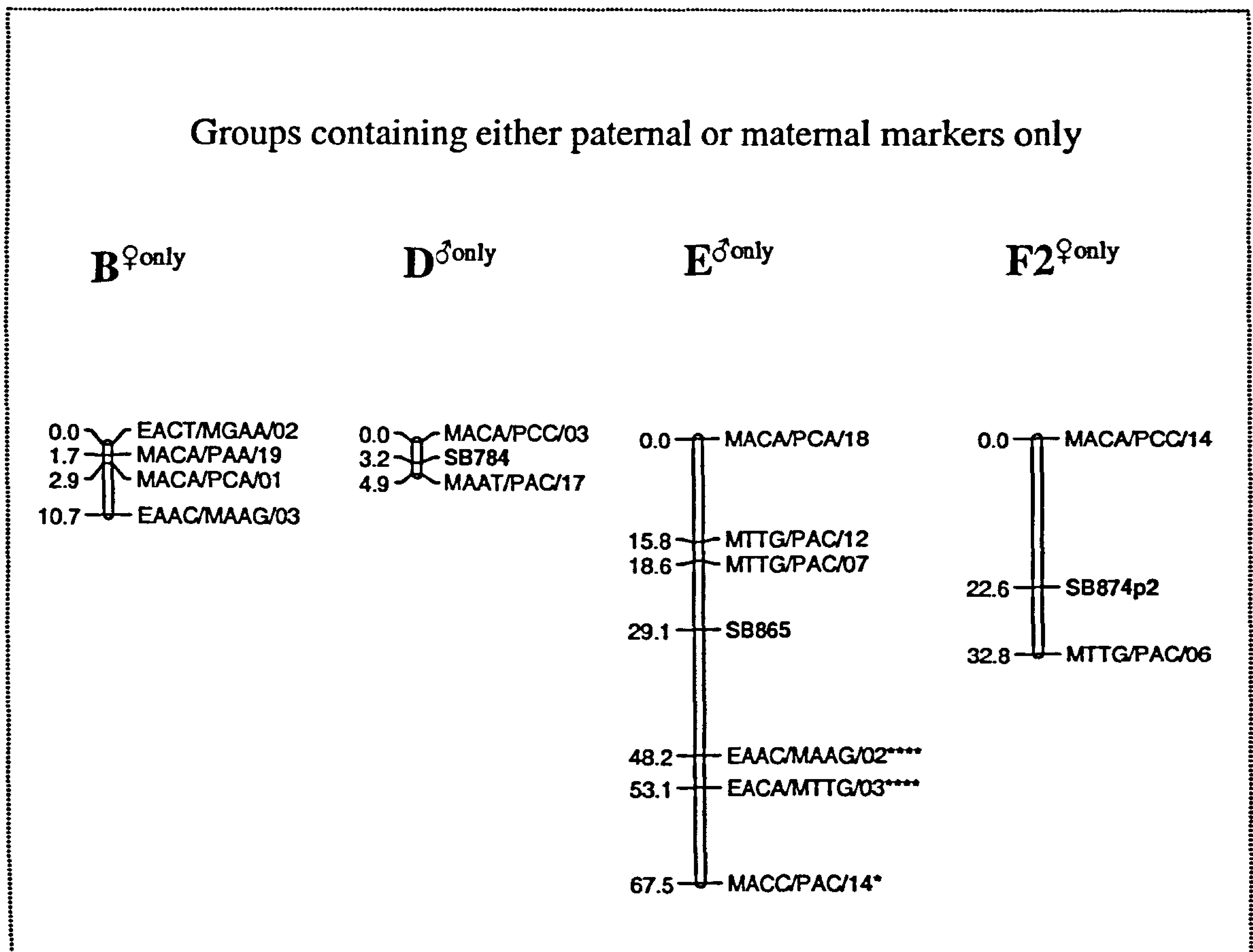


Figure 3.8 continued. The K3 consensus linkage map.

3.4 Discussion

The construction of a genetic map of *S. viminalis,* a willow species commercially important for biomass production, is of fundamental importance to future molecular breeding endeavours. At the time that construction of the K3 map was initiated, no linkage maps for any willow species were available. However, mapping efforts in Sweden have since resulted in the publication of an AFLP/RFLP-based linkage map of a cross between the biomass cultivar *S. viminalis* var. Bjorn and *S. schwerinii* clone "L78183" (Tsarouhas *et al.*, 2002). In contrast, a number of genetic maps have previously been reported for poplar species (Bradshaw *et al.*, 1994; Wu *et al.*, 2000; Cervera *et al.*, 2001). As poplars are closely related to willows, opportunities should now arise for comparative genetic analysis between these important tree species.

3.4.1 Choice of molecular markers

The AFLP technique was chosen as it generates a large number of polymorphic markers per single assay (Vos and Kuiper, 1997) and, therefore, has the capacity to generate a large number of markers in a relatively short timeframe. Furthermore, AFLP markers are reliable and easily reproduced, in contrast to other PCR-based dominant marker systems, such as RAPDs, where reproducibility between laboratories has proven difficult (Jones *et al.*, 1997). However, in certain configurations, AFLP markers may provide limited information for estimation of recombination frequencies and hence, may not be the marker of choice when mapping in outbreeders. For this reason, microsatellites were also employed. Although costly and time-consuming to develop, these co-dominant markers are highly informative in linkage analysis, particularly in *abxac* or *abxcd* configurations and are useful for anchoring maps derived from different populations. In addition, the PCR-based nature of microsatellites makes them more amenable to automation and high-throughput screening than hybridisation-based RFLP markers (also co-dominant).

3.4.2 AFLP markers

Both AFLP techniques detected a large number of polymorphic loci, which is commensurate with the outbreeding/heterozygous nature of this species (Barker *et al.*, 1999). In general, fewer polymorphic AFLP markers were produced using the two-step PCR method of Vos *et al.* (1995) compared with the magnetic bead selection method (Zabeau and Vos, 1993). This difference may be attributed to a number of factors, including differences in the restriction enzymes used and the inclusion, in the two-step protocol, of an additional selective nucleotide on the *Eco*RI primer used in the selective amplification. However, the lower polymorphism obtained with this protocol was outweighed by an increase in throughput and a reduction in consumption of genomic DNA. Two restriction enzyme combinations, *MseI/PstI* and *MseI/Eco*RI, differing in sensitivity to cytosine methylation, were used in an attempt to achieve a wider distribution of markers across the genome, on the basis of the observation that sensitivity to cytosine methylation influenced map positions of AFLP markers in soybean (Young *et al.* 1999).

3.4.3 Willow microsatellite markers

The development of a large number of willow microsatellite markers during this study was facilitated by the highly enriched nature of the microsatellite library produced by Edwards *et al.* (1996). The failure of a several primer sets to generate PCR products in the K3 pedigree may be a consequence of nucleotide variation within the selected priming sites of the *S. viminalis* population under study and the species (*S. burjatica*) from which the library was made. Alternatively, this may be attributed to the use of sub-optimal PCR conditions. The lack of amplification could potentially be circumvented by the selection of alternative priming sites or by more thorough optimisation of PCR conditions. These approaches were not explored given that the time available for marker development was limited and that the availability of suitable library clones was not a limiting factor in the production of alternative markers. Furthermore, for use in subsequent mapping studies, in which PCR multiplexing would be used (Chapter 5), it was preferable to develop a panel of microsatellite markers that would successfully amplify target loci under the same PCR conditions.

While fragment detection based on [F]dCTP labelling was successfully used to map microsatellite loci SB525 and SB565, such an approach is not applicable to resolution of alleles that are similar in size (less than 4 bp difference) (Applied Biosystems; personal communication) and could not, therefore, be used to map all available microsatellite loci. Furthermore, at the time when microsatellite mapping was being performed, access to the ABI377 DNA Sequencer at LARS was severely limited due to heavy usage by other research groups. For these reasons, this specific use of fluorescence was not used further in this project.

The development of willow microsatellite marker during this mapping study provides an important resource for future genetic studies in this genus. Microsatellites can be used to anchor maps generated from different crosses. Moreover, they can be detected using simple assays that can be adapted for high throughput screening (Mazur and Tingey, 1995). This can be very useful if microsatellite markers are found closely linked to agronomic traits of importance since they can be easily employed in MAS (Section 1.4.5). Microsatellites are also very informative in the characterisation of natural populations and germplasm collections, particularly when their map positions are known (Karp *et al.*, 1996). Some of the microsatellites described in this and later chapters are currently being used for the genetic characterisation of the National Willow Collection at LARS and in populations studies of *S.fragilis*, *S. alba* and their hybrids. Ultimately, microsatellites could also be used for variety identification in willow through the construction of a database of variety "fingerprints".

3.4.4 Poplar microsatellites

Synteny between genomes (in which the gene order of genomes is conserved but not necessarily gene function) is now a well described phenomenon (Section 1.4.4). Given their close taxonomic relationship, it may be considered likely that poplar and willow genomes possess significant levels of synteny and collinearity. The exchange of sequence-based molecular markers, such as microsatellites, between poplar and willow has the potential to improve the efficiency of genetic studies in both genera, and may provide a basis for comparative mapping and QTL studies. Furthermore, comparative genetic approaches will become increasingly important, especially once the complete annotated poplar genome sequence becomes available ((6x draft sequence expected by the end of August 2003; DOE Joint Genome Institute, USA).

In this study, a proportion of microsatellite markers (5 out of 19) developed for use in poplar successfully amplified willow DNA. However, it has been suggested more recently that up to 75% of poplar microsatellites identified as a part of the current poplar genome sequencing effort, may be useful in willow (G.A. Tuskan, personal communication). Although incomplete for *S. burjatica*, "Germany" and the poplar hybrid "Ghoy", sequence data obtained for microsatellite locus PMGC223 was sufficient to suggest that corresponding loci were amplified in willow and poplar. Generation of complete sequence data for this locus, which could have been achieved by cloning PCR products prior to sequencing, was not pursued further during this project. Failure of a number of poplar primer sets to amplify willow DNA could be attributed to sub-optimal PCR conditions, although the difference in the sequence of regions flanking the microsatellite suggests that prevention of specific primer annealing may be a more probable cause. The two mapped poplar microsatellites (PMGC93 and PMGC2020) showed linkage on Group II of the willow map although they were not reported to be linked in poplar. However, PMGC2020 mapped to a different group in the *P. trichocarpa* map compared to the *P. deltoides* and *P. nigra* maps (Cervera *et al.*, 2001).

3.4.5 ESTP markers

The efficacy of developing co-dominant willow ESTP markers based on expressed sequences from *Populus* for use in *Salix* was confirmed, although, due to time limitations, expression of these sequences in the latter was not investigated, nor were willow products sequenced for comparison with the original ESTs. Both of these aspects should be studied more comprehensively in future. Of the two loci tested, one possessed a significant size polymorphism and could be mapped using the visualisation method employed (agarose gel electrophoresis). While the detection of size polymorphisms using ESTP markers has been reported (Temesgen *et al.*, 2001), it is likely that, given the greater degree of conservation in coding regions compared to non-coding regions (Cato *et al.*, 2001), the detection of polymorphisms of this type may be the exception rather than the rule. Therefore, alternative polymorphism detection techniques, such as SSCP, PCR-RFLP or analysis of SNPs may be more suitable for efficient ESTP mapping in future.

Results obtained here suggest that the growing amount of *Populus* EST data may be exploited in the development of sequence-based co-dominant markers for use in *Salix*. The generation of additional markers of this type may provide a more efficient route towards comparative mapping between poplar and willow than that afforded by markers targeting non-coding genomic regions, i.e. microsatellites (Section 1.4.4).

3.4.6 Segregation analysis

The identification of markers displaying distorted segregation ratios has previously been described for a number of tree species (Bradshaw and Stettler, 1994; Cai *et al.*, 1994; Grattapaglia and Sederoff, 1994; Lanaud *et al.*, 1995; Krutovskii *et al.*, 1998; Marques *et al.*, 1998; Arcade *et al.*, 2000; Tsarouhas *et al.*, 2002). The percentage of distorted markers detected here (18%) is equivalent to that reported in the willow map of Tsarouhas *et al.* (2002), and is similar to levels observed in linkage analyses of pine (14-15%; Kubisiak *et al.*, 1995) and oak (18%; Barreneche *et al.*, 1998). While segregation distortion in itself is not a reason to exclude markers from mapping studies, AFLP markers with highly skewed segregation ratios may represent superimposed loci of equivalent electrophoretic mobility that segregate independently in the mapping population and, if included in analyses, may result in erroneous maps. For this reason, highly distorted AFLP markers identified in this study were discarded. Furthermore, if a marker is distorted because it is linked to a gene involved in viability or sex-linkage for example then, due to linkage alone, all markers in that region should display distorted segregation ratios. Highly distorted markers that do not map to such regions should therefore be treated with caution. However, it is important that not all distorted markers are excluded from mapping studies as such markers can highlight regions of interest. For example, in a mapping study of poplar, Cervera *et al.* (2001) observed deviant segregation ratios in markers known to co-segregate with the gene for resistance to *Melampsora* rusts and suggested that the reason for distorted segregation ratios was the death of susceptible trees. Had these markers been excluded from the analyses, a functionally important region of the linkage map would have been omitted. Clusters of distorted markers have been observed in a number of tree species including poplar (Cervera *et al.*, 2001), eucalyptus (Verhaegan and Plomion, 1996), pine (Kubisiak *et al.*, 1995), oak (Barreneche *et al.*, 1998) and now willow.

3.4.7 The K3 parental maps

For construction of each parental map, AFLP and microsatellite markers that segregated only in the respective parent and, in addition, parental alleles from microsatellites of type *abxac* and *abxcd*, were used. The inclusion of markers segregating in both parents results in the estimation of recombination frequency for heterozygous marker pairs averaged for both male and female meioses and may, therefore, differ from parental estimates. This can give rise to conflicting marker orders between parental and consensus maps. For this reason, it was preferable to first construct independent parental maps to allow comparison of marker orders with those determined for a consensus map.

For both the paternal and maternal maps, the number of resolved linkage groups exceeded the haploid chromosome number of willow (n=19) suggesting that additional markers will be required to bridge gaps between linkage groups and improve resolution. The smaller total map length of the paternal map (775.2 cM) in comparison to the maternal map (910.2 cM) suggests that a number of regions are still under-represented by those markers available in the paternal data set. For example, in group I of the consensus map (Figure 3.8) only three paternal markers showed linkage to this group. In the paternal map, one of these markers failed to link to any group, however, linkage could be established for the two other markers and the resulting marker pair spanned only 14 cM. In contrast, the corresponding linkage groups of the maternal map and consensus maps spanned 49.1 and 96.9 cM, respectively.

3.4.8 The K3 consensus map

As with the parental maps, the number of resolved linkage groups exceeded the willow haploid chromosome number. The incorporation of multiple groups of the parental maps into single consensus groups I and II suggests that linkage group resolution may be improved with additional markers in the future. The inclusion of AFLP marker data of type *a0xa0* in construction of the consensus map resulted in the identification of additional linkage groups that were not represented in the parental maps. This observation was not unexpected given that the parents of the mapping population were full-sibs and therefore, large regions of the resulting K3 map would be expected to comprise only AFLP markers of this class. Problems encountered when marker data

of this type were included in the analysis were highlighted by occasional discrepancies in marker orders between parental and consensus maps. In linkage analysis of outbred populations, *a0xa0* markers can be extremely uninformative in certain configurations and very often lead to estimation frequency estimates of 0.0 (Maliepaard *et al.*, 1997). A number of markers were thus excluded in construction of the consensus map in order to achieve co-alignment with homologous linkage groups of the parental maps. However, inclusion of this marker class is beneficial for determination of linkage between maternal and paternal markers (types *aaxab* and *abxaa*) which cannot be established directly. This was exemplified by groups Am and Af, which showed linkage in the presence of two *a0xa0* AFLP markers in the consensus map data set, although these did not provide sufficient linkage information to allow map construction for this group as a whole. The identification of so-called suspect recombination estimates for a number of marker pairs may be a consequence of multiple homeologous loci existing in the *S. viminalis* genome. For example, if two homeologous segments of the same AFLP fragment were amplified, being heterozygous in one parent at both loci and homozygous absent at both loci in the other, a band would only be observed in one parent but would be over-represented in the full-sib offspring. Such scenarios would influence recombination estimates and result in suspect linkages.

The distribution of AFLP markers across all linkage groups of the willow map indicates that this technology was particularly useful for rapid map coverage. The identification of obvious AFLP clusters on several groups is not uncommon and has been observed in previously published genetic maps including poplar (Wu *et al.*, 2000), rose (Debener and Mattiesch, 1999) and soybean (Young *et al.*, 1999). This phenomenon may be explained by reduced recombination in chromosomal regions such as centromeres (Tanksley *et al.*, 1992; AlonsoBlanco *et al.*, 1998). Microsatellite markers were located on 18 of the 28 linkage groups identified[†] and also showed good distribution across the map. These sequence-based markers will prove extremely useful for identification of homologous linkage groups in future mapping studies with new populations aimed at QTL analysis.

[†] linkage groups of the consensus map and groups exclusive to either parental map

By employing relatively stringent parameters for analysis, and by comparison of marker orders and map distances with parental maps, a statistically robust consensus map of willow was constructed. This genetic map was used as a framework for further mapping efforts described in Chapter 5, and also in QTL analysis (Chapter 6).

3.5 Chapter summary

- A genetic linkage map of *S. viminalis* was constructed based on the K3 mapping population according to the double pseudo-testcross approach.
- Significant investment into marker development was required, ultimately resulting in the generation of 56 polymorphic microsatellites and 477 AFLP markers for inclusion in linkage analysis.
- 19 microsatellite primer sets originally developed for poplar were tested in willow. Five loci were successfully amplified in the mapping pedigree, of which two segregated and were mapped.
- Two publicly available *Populus* ESTs were used as a basis to test the efficacy of developing ESTP markers for use in willow. Primers designed to these two sequences resulted in amplification of willow products. One locus was polymorphic in the mapping pedigree and was subsequently mapped.
- The paternal map comprised a total of 132 AFLP markers and 25 microsatellite markers that mapped to 21 linkage groups of three markers or more. Resulting groups spanned 777.5 cM with an average marker interval of 7.9 cM.
- The maternal map comprised 139 AFLP markers and 33 microsatellite markers located on 25 linkage groups of three markers or more. These groups spanned 910.2 cM with an average interval between markers of 8 cM.
- The resulting 21 K3 consensus groups[‡] comprised 191 AFLP markers of types *aaxab* and *abxaa*, 81 AFLP markers of type *a0xa0*, one ESTP marker and 43 microsatellites of all possible segregation types.
- This genetic map will be used as a framework for mapping in population K8 as described in Chapter 5, and also as in QTL analysis (Chapter 6).

[‡] defined here as groups previously identified in the parental maps but comprising at least one marker heterozygous in both parents

Chapter 4. Selection and establishment of a large segregating population as a basis for QTL mapping

4.1 INTRODUCTION 117

4.1.1 Selection of candidate populations 118

4.1.2 Increasing the number of K3 progeny 122

4.1.3 Experimental design 122

4.1.4 Pedigree authentification 125

4.2 MATERIALS AND METHODS 125

4.2.1 DNA extractions for pedigree authentification 125

4.2.2 AFLP analysis of candidate pedigrees 125

4.2.3 Pedigrees of the chosen crosses K1 and K8 126

4.2.4 Preparation of cuttings for planting the field trial 127

4.2.5 Planting the field trial 128

4.2.6 Field-based trait assessments 128

4.2.6.1 Field-based assessments of rust resistance 129

4.2.6.2 Destructive yield assessments 129

4.2.6.3 Stem diameter assessments 130

4.2.6.4 Stem height assessments 130

4.2.7 Laboratory-based trait assessments 130

4.2.7.1 Laboratory-based assessment of rust resistance 131

4.2.7.2 Laboratory-based beetle feeding experiments 132

4.3 RESULTS 135

4.3.1 AFLP analysis for authentification of parents used in candidate pedigrees 135

4.3.2 Establishment of the field trial 137

4.3.3 Trait assessments 138

4.3.4 Laboratory-based trait assessments 142

4.3.4.1 Laboratory-based assessment of rust resistance 142

4.3.4.2 Laboratory-based beetle feeding experiments 144

4.4 DISCUSSION 145

4.4.1 AFLP analysis of candidate pedigrees 145

4.4.2 Field-based trait assessments 146

4.4.2.1 Field-based assessment of rust resistance 146

4.4.2.2 Assessments relating to yield 148

4.4.3 Laboratory-based assessments 148

4.4.3.1 Laboratory-based assessment of rust resistance 149

4.4.3.2 Laboratory-based beetle feeding experiments 149

4.5 CHAPTER SUMMARY 151

4.1 Introduction

This Chapter describes the choosing and establishment of a large segregating mapping population for use in QTL analysis aimed at identifying genomic regions underlying traits of agronomic importance. The rationale for selecting a mapping population to use as a basis for further study and the experimental procedures leading to this choice, are discussed.

There were two main reasons for establishing a new mapping population. First, the K3 family, used for construction of the genetic map (Chapter 3) comprised a relatively small number of progeny. The size of mapping population used for QTL analysis can, depending on the heritability of the traits under study (Moreau *et al.*, 1999), greatly influence the outcome of such studies in several ways. For example, simulation studies by Beavis (1994) suggested that in QTL mapping experiments based on populations of 100 - 200 progeny, only a fraction of the true QTL may be detected. Furthermore, this study predicted that estimates of phenotypic effect attributed to each locus would generally be exaggerated with populations of this size. In order to gain more comprehensive and accurate information regarding QTL, studies based on populations comprising at least 500 progeny were recommended. The predictions made in this simulation study were later confirmed experimentally by comparison of results obtained in QTL analysis based on two maize testcross populations, both derived from the same parents but differing in population size (Melchinger *et al.*, 1998). In this experiment, a greater number of QTL were detected in the larger of the two populations.

In addition to increasing the detection power of QTL mapping experiments, the use of large population sizes can also increase the precision with which QTL can be mapped. Often, depending on the heritability of the trait, confidence intervals associated with QTL locations can be large (Van Ooijen, 1992; Darvasi *et al.*, 1993; Hyne *et al.*, 1995). These can be difficult to reduce, even if local marker density is improved (Kearsey and Farquhar, 1998). Increasing the number of genotypes included in a QTL study can be an efficient method of improving mapping precision (Asíns, 2002).

Second, although additional K3 progeny could have been generated from a further mating between the parents, previous field-based observations suggested that the level

of phenotypic variation for all traits to be targeted might not have been sufficient to achieve the goals of this study. For example, data resulting from field-based assessments of rust resistance/susceptibility performed in Sweden suggested that all K3 progeny were susceptible to rust infection, although the severity of resulting disease was shown to vary between individuals (I. Åhman, personal communication). These results may suggest that potentially important genes conferring resistance to rust are not segregating in this population, making K3 an inferior choice for use as the mapping population underpinning this present study. Furthermore, as the pedigree of the K3 cross comprised only pure *S. viminalis* species, it was considered unlikely that sufficient variation in resistance/susceptibility to willow beetle herbivory would be present for identification of underlying QTL (Section 1.2.3.2.1).

For these reasons, a new willow mapping population, more suited to QTL analysis, both in terms of population size and segregation of important agronomic traits, was established.

4.1.1 Selection of candidate populations

From 18 diploid crosses originally produced from parental material in the National Willow Collection (NWC), as part of the European Willow Breeding Programme (EWBP), seven were highlighted as potential candidate families for use in this study (Table 4.1). These crosses had all produced sufficient numbers of seed to allow for the establishment of a large population for subsequent QTL analysis. Furthermore, based on field assessments performed by the willow breeder at LARS, the parental pedigrees suggested that the traits being targeted in this present study had a high chance of segregating in the various progenies (K.N. Lindegaard, personal communication). However, clearly this could not be guaranteed for all crosses or for all traits of interest. For this reason, two populations were initially chosen for inclusion in the field trial. From these, the choice would be narrowed to a single population for use in QTL mapping on the basis of results obtained in trait assessments performed during the establishment year.

Prior to making the final choice of the two candidate mapping populations, molecular marker analysis of the expected pedigrees of each of the candidate crosses was

performed (Section 4.2.2). These experiments led to the exclusion of three of the seven candidate crosses (Section 4.3.1).

As resistance to rust diseases is of key importance in the success of SRC plantations (Section 1.2.3.1.1), the final choice of most suitable crosses was made largely on the basis of field-based rust assessments performed at LARS and in Sweden as part of the willow breeding scheme. Taking all the above into account, the progenies of crosses K1 and K8 were selected as the two mapping populations that were then established in the field trial for use in subsequent analyses.

The statistical design of the field trial was developed in consultation with the Long Ashton statistician Dr. Phil Brain and also Mr. Guy Donaldson, the farm manager and the Field Experiments Committee at Long Ashton who have considerable experience of designing willow field trials.

Table 4.1. Details of the seven willow crosses originally proposed as candidate mapping populations.

Cross	Female Parent	Male Parent	Maternal Grandparents		Paternal Grandparents	
			♀	♂	♀	♂
524	LA940112 *S.viminalis x S.triandra*	LA940143 *S.viminalis x S.triandra*	115/34 "Bowles Hybrid" *S. viminalis*	112/18 "Brunette Noire" *S. triandra*	115/34 "Bowles Hybrid" *S. viminalis*	112/18 "Brunette Noire" *S. triandra*
529	115/34 "Bowles Hybrid" *S.viminalis*	LA940143 *S.viminalis x S.triandra*	Unknown *S. viminalis*	unknown *S. viminalis*	115/34 "Bowles Hybrid" *S. viminalis*	112/18 "Brunette Noire" *S. triandra*
553	LA940140 *S.viminalis x S.triandra*	LA940143 *S.viminalis x S.triandra*	115/34 "Bowles Hybrid" *S. viminalis*	112/18 "Brunette Noire" *S. triandra*	115/34 "Bowles Hybrid" *S. viminalis*	112/18 "Brunette Noire" *S. triandra*
581	LA960230 *S.viminalis x S. schwerinii x S. burjatica*	LA960231 *S.viminalis x S. schwerinii x S. burjatica*	33/08 "Pavainen" *S. burjatica*	SW910006 (var. Björn) *S. viminalis x S. schwerinii*	33/08 "Pavainen" *S. burjatica*	SW910006 (var. Björn) *S. viminalis x S. schwerinii*
K0	K477-S3 *S. viminalis x S. schwerinii*	K477-R11 *S. viminalis x S. schwerinii*	SW880435 (var. Astrid) *S. viminalis*	SW930984 *S. viminalis x S. schwerinii*	SW880435 (var. Astrid) *S. viminalis*	SW930984 *S. viminalis x S. schwerinii*
K1	K477-R6 *S. viminalis x S. schwerinii*	K477-R11 *S. viminalis x S. schwerinii*	SW880435 (var. Astrid) *S. viminalis*	SW930984 *S. viminalis x S. schwerinii*	SW880435 (var. Astrid) *S. viminalis*	SW930984 *S. viminalis x S. schwerinii*
K8	K477-S1 *S. viminalis x S. schwerinii*	K477-R11 *S. viminalis x S. schwerinii*	SW880435 (var. Astrid) *S. viminalis*	SW930984 *S. viminalis x S. schwerinii*	SW880435 (var. Astrid) *S. viminalis*	SW930984 *S. viminalis x S. schwerinii*

The parents of both the K1 and K8 crosses were full-sib members of the K477 population, originally produced by Svalöf Weibull AB in Sweden. Of the original population, 21 individuals were retained by the breeding company following initial selections on the basis of either high or low levels of rust resistance. Of the remaining progeny, 14 were originally considered to be broadly resistant to rust and were labelled with the prefix 'R' (R1 - R14). The remaining seven family members were found to be more susceptible to infection and were labelled with the prefix 'S' (S1 – S7). It is important to note that these prefixes were assigned on the basis of original rust assessments performed in Sweden and were relative, i.e. rust resistance was not necessarily complete for all individuals labelled as resistant and the levels of susceptibility varied from one susceptible individual to another (Table 4.2).

Table 4.2. K477 population field-based rust scores recorded in 1996 and 1997 in Sweden. Scores were assigned on a scale ranging from 0 – 6 where 0 = no rust found: 6 = severe rust infection (Åhman, 1997). Mean values derived from assessment of three clonal replicates per K477 individual are provided. Parental (SW880435 var. Astrid and SW930984) and grandparental (SW880435 var. Astrid and SW910006 var. Björn) scores are also shown.

Progeny	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14
Rust Score 1996	0.5	1	0.33	0	1	0.17	0	0.17	0.17	2.17	0.33	0.17	0.17	0.17
Rust score 1997	0.5	0	0	0	1.5	0.17	0.17	0.17	0.17	0.17	0	0	0	0
Mean Rust Score	0.5	0.5	0.17	0	1.3	0.17	0.09	0.17	0.17	1.17	0.17	0.09	0.09	0.09

	S1	S2	S3	S4	S5	S6	S7	Astrid	930984	Björn
Rust Score 1996	1.7	2	3	3	2	2.67	3.67	3	0.17	0
Rust score 1997	2.2	2	2.83	3	2.7	3.5	4.5	3.83	0	0
Mean Rust Score	1.9	2	2.92	3	2.3	3.09	4.09	3.42	0.09	0

Source: I. Åhman, Svalöf Weibull AB, Sweden.

The two crosses selected for further study, K1 and K8, were of types R x R and R x S, respectively (Figure 4.2). As the genetic basis of resistance was unknown at this time, it was important to select these two contrasting crosses to increase the likelihood that rust resistance would segregate in at least one mapping population. To illustrate, in the hypothetical case that a single dominant gene confers complete resistance, then two scenarios in which segregation would not occur were possible. First, if in the R x R

cross if at least one parent was homozygous resistant (RR) then all offspring would be of type RR or Rr and no phenotypic segregation for rust resistance would occur. Similarly, in the case of the R x S cross, if one parent was homozygous resistant (RR) and the second parent was homozygous susceptible (rr), then all progeny would be of genotype Rr and no segregation for rust phenotype would be observed.

4.1.2 Increasing the number of K3 progeny

As the segregation of traits of interest could not be guaranteed for populations K1 and K8, a further cross between the K3 parents was performed to generate a greater number of progeny as a contingency for use in QTL mapping should the K1 and K8 populations prove to be uninformative. The increased progeny size would increase the power of any future QTL studies based on this population (section 4.1). An additional 143 seeds were produced and one year old stems were grown and used to make cuttings for inclusion in the field trial, increasing the number of K3 full-sibs to 209.

4.1.3 Experimental design

After the K1 and K8 segregating populations had been chosen as candidate mapping populations for use in this study, careful consideration was given to the design of the field trial from which trait data would be generated. This was of fundamental importance to the power of the experiment, both in terms of the ability to detect QTL and also the robustness of associated results.

The first major consideration concerned the question of whether to plant several replicates of each genotype at different positions within the trial site. By assessing traits of interest at various Jorr positions within the site, the effects of environmental variation can be accounted for, thus, the heritability of individual QTL may be enhanced and the power of the QTL study improved (Kearsey and Farquhar, 1998). However, the K1 and K8 populations had only been grown for a single season in the nursery and the amount of material available for the production of cuttings was limited. In order to generate sufficient material to implement such a planting regime, a further year's growth would have been necessary. Due to time limitations imposed on the project, this was not a feasible option and an alternative strategy was adopted, in which, six clonal replicates of each K1 and K8 genotype were planted as single plots,

each in a 2 x 3 arrangement, within the field experiment (Figure 4.1c). To facilitate identification of possible environmental inconsistencies across the trial site, and to allow subsequent adjustment of trait values prior to QTL analyses (Section 6.2.2), a reference willow variety was planted at several pre-selected plot positions throughout the site (Figure 4.1). This provided a robust control of spatial heterogeneity as all such plots would contain genetically identical individuals and any variation identified between them would be a consequence of environmental influences only. The biomass cultivar, *S. viminalis* var. Jorr, was selected as the reference variety for several reasons. First, as a large amount of planting material would be required, it was important to select a genotype for which sufficient material was available. For this reason, a commercially available cultivar was selected. Second, as the mapping populations were not expected to yield as well as some of the elite biomass varieties, the possibility of extensive competition between any reference plots and nearby plots comprising progeny of mapping population was considered. The variety Jorr generates a low to mid-range yield in comparison to some of the more recently marketed cultivars and was, therefore, deemed suitable for this role. Furthermore, this variety is susceptible to both rust disease and willow beetle herbivory and would not, therefore, be expected to act as a barrier to the spread of either of these pressures throughout the field trial.

The cultivar Jorr was also planted in a double buffer row around the perimeter of the mapping populations to minimise any 'edge effects' that may influence phenotype. For example, in the absence of any buffer row, the growth of mapping population individuals located at the edge of the trial may have been more heavily influenced by environmental factors such as increased exposure to harsh weather conditions, decreased competition from neighbouring plots, greater light availability, etc. In addition, it was necessary to include seven double tramline rows within the population to allow for tractor access.

In order to keep the mapping study relevant to current SRC growing practices, the spatial design of the field trial was based on planting regimes similar to those employed in commercial biomass plantations (Figure 4.1c). Furthermore, the trial would be harvested in accordance with a standard SRC rotation scheme, i.e. the trial would be cut back following the establishment year and again after the third year of growth. The final design of the field experiment is shown in Figure 4.1.

Figure 4.1. The final field trial plan comprising the K1, K8 and K3 mapping populations. The plots comprising six replicates of each of the K1, K8 and K3 genotypes are labelled 1001-1946, 8001-8947 and 3001-3209, respectively. Shaded boxes labelled Jorr indicate plots comprising *S. viminalis* var. Jorr for use as a control of spatial heterogeneity across the site. Additional shaded areas represent the planting of Jorr as buffer or tramline rows. Figure 4.1a shows the entire field plan, Figure 4.1b shows a magnified view of a section of the field trial and Figure 4.1c illustrates the planting arrangement within and between plots (not to scale).

4.1.4 Pedigree authentification

The pedigree of individuals used as parents for the candidate mapping populations (Table 4.1) were checked using molecular markers by screening their grandparents and parents in the NWC (see Section 1.2.4.1). Only crosses where the lineage had no discrepancies were chosen. As the work progressed, however, and the progeny of the crosses were established, discrepancies in marker profiles between progeny and parents became evident, suggesting that the identification of the parents from individuals R1-R14 and S1-S7 of K477 may not have been accurate (Section 5.1.2). These data were not known at the time of planting the populations but clearly has subsequent implications for the work described in this Chapter. In particular, to a large extent, it meant that the trait information generated for the expected parents as part of the work described in this Chapter was not directly relevant to the study. However, although not discussed in detail here, the incorrect parents are included in descriptions of methods and results to illustrate their consideration in experimental designs. The identification of the correct parents for the selected cross is described in Chapter 5. The implications of the inclusion of the incorrect parents in early stages of this study are discussed in greater detail in Chapters 5 and 6.

4.2 Materials and Methods

4.2.1 DNA extractions for pedigree authentification

For DNA extractions, leaf material was collected from parents and grandparents growing as part of the NWC at LARS. The Nucleon Phytopure Plant DNA Extraction Kit was used for all extractions as described in Section 2.4.1. Resulting DNA quality and quantity was determined by agarose gel electrophoresis (Section 2.4.2).

4.2.2 AFLP analysis of candidate pedigrees

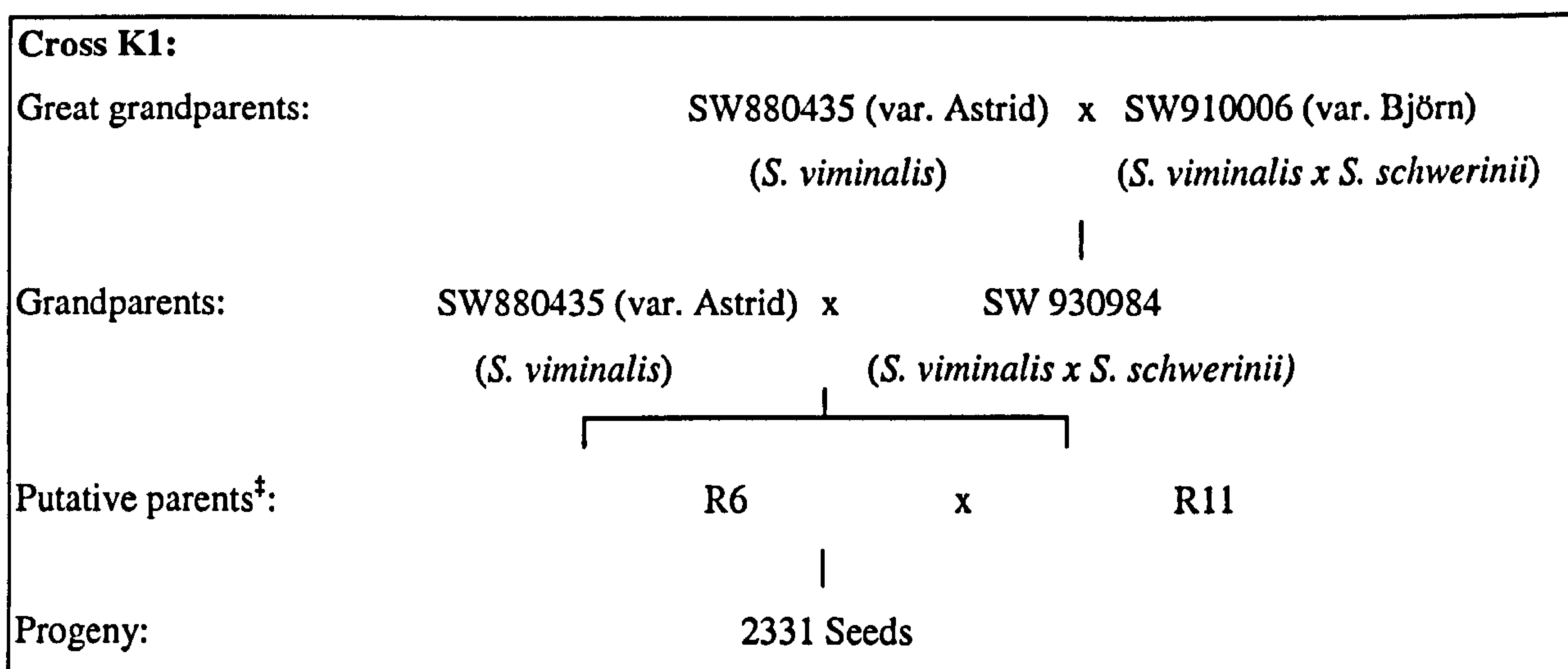
AFLP analysis was performed using two primer combinations (MACA/PCC and MACA/PAC) according to AFLP Method 1 (Section 2.5.2.1). Profiles for all parents and grandparents of the candidate crosses were generated (individuals are listed in Table 4.1). To test the robustness of the technique, two DNA samples extracted in two independent years were included for varieties Björn and Astrid. To confirm that the

expected pedigrees of the potential crosses were possible, resulting autoradiographs were examined to confirm that all AFLP bands found in the candidate parent profiles were also observed in the profile of a respective grandparent.

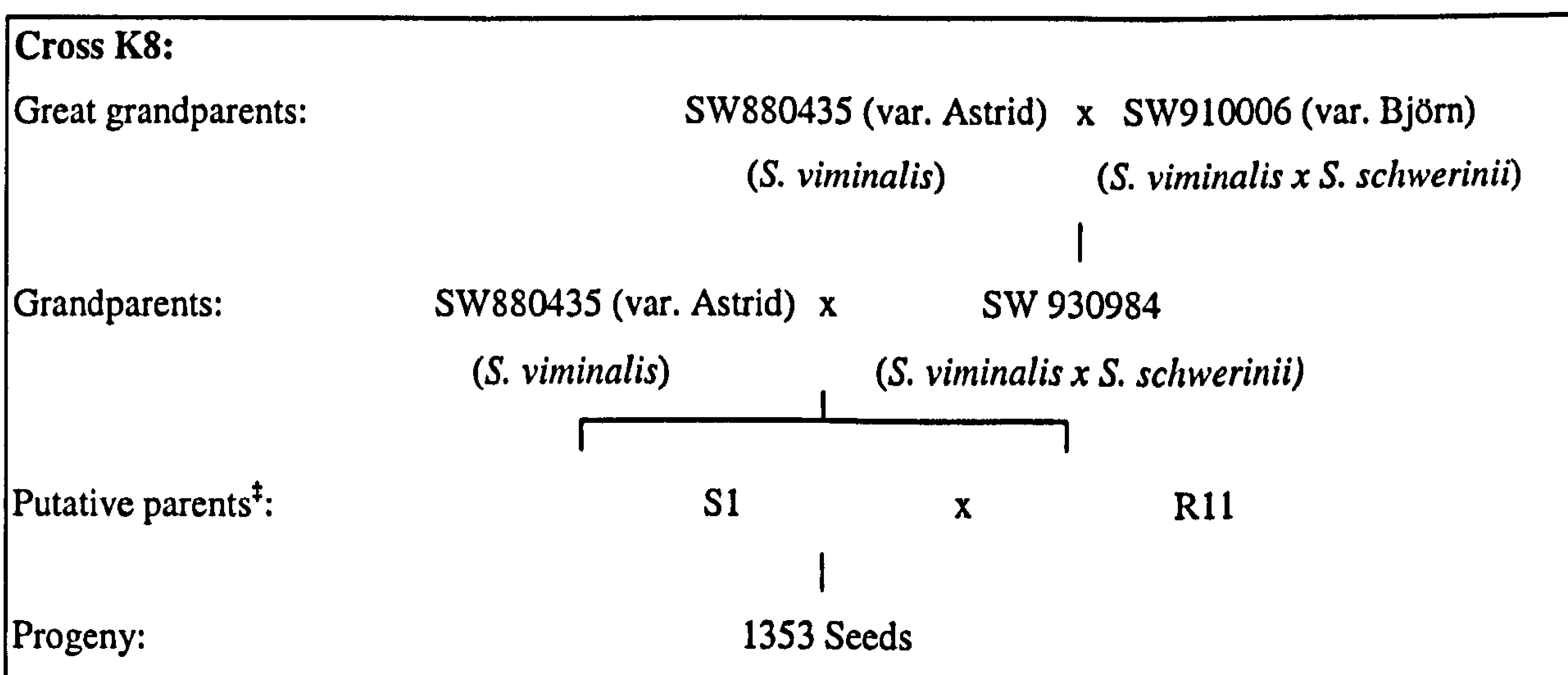
4.2.3 Pedigrees of the chosen crosses K1 and K8

The expected pedigree information of the two selected crosses for inclusion in the field trial is shown in Figure 4.2. All plant material used to make the crosses was collected from the NWC at LARS. Crosses were performed as described in Lindegaard and Barker (1997) as part of the European Willow Breeding Programme.

Figure 4.2. Expected pedigrees of the K1 and K8 crosses.



‡ see Section 5.1.2

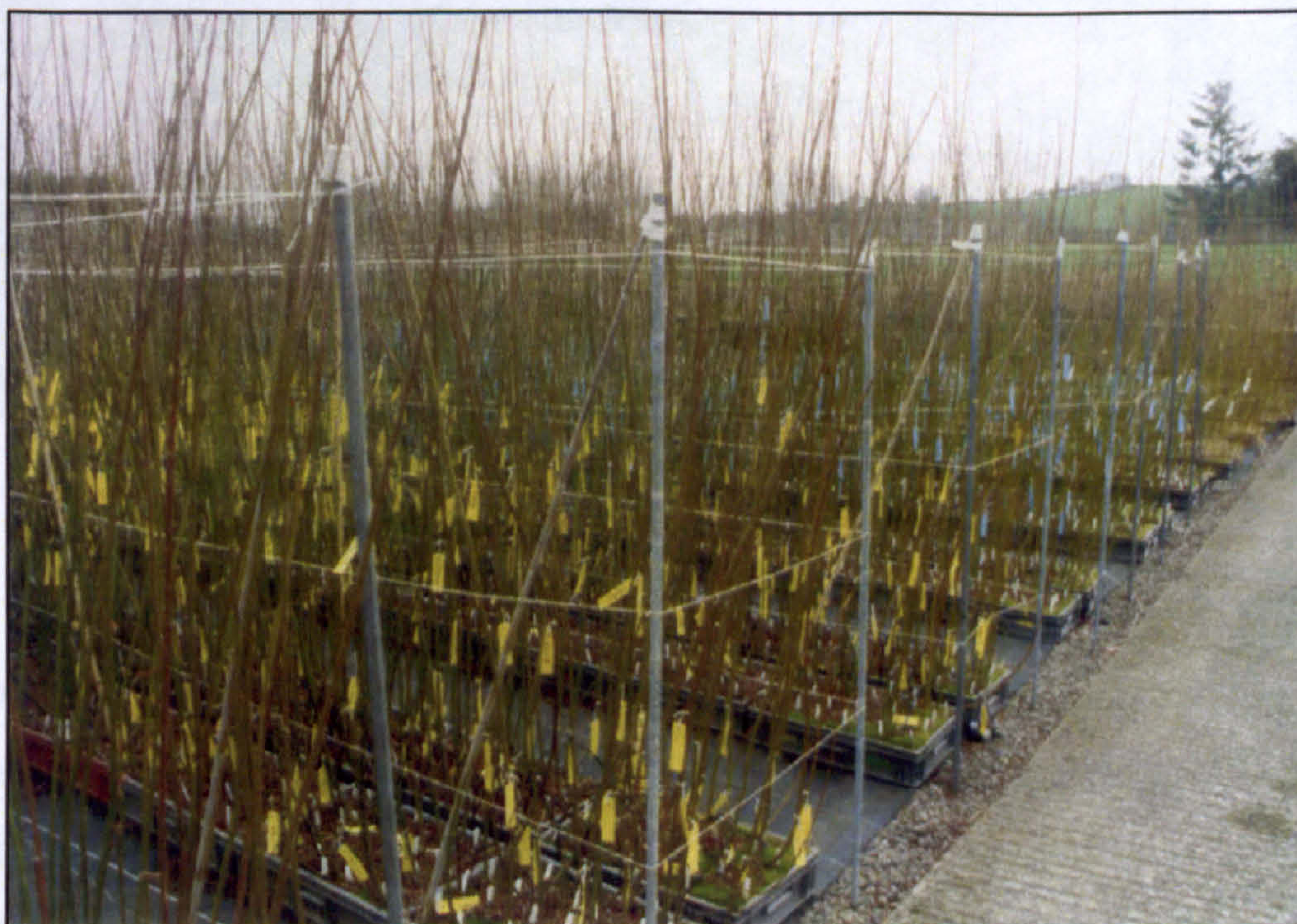


‡ : see Section 5.1.2

4.2.4 Preparation of cuttings for planting the field trial

After the choice of crosses for inclusion in the field trial had been made (K1 and K8), 950 seeds from each cross were germinated and grown in a glasshouse. When sufficient growth was obtained, seedlings were pricked out into trays, each containing 24 plants, and transferred to the nursery. Following one season of growth, viable progeny from each cross were individually labelled (Figure 4.3) and stem heights were recorded. All individuals were then harvested and each used to generate six 15 cm cuttings per genotype for planting in the field trial. Where the length of an individual was too short for the production of six cuttings, rather than make six cuttings of shorter length, the maximum number of 15 cm cuttings possible was made. Cuttings of the 'reference' clone used for planting throughout the trial, within the mapping populations, were generated from the biomass variety 'Jorr' taken from the NWC. Additional 'Jorr' material for use as buffer rows and tramlines was purchased from Murray Carter. Parental and grandparental material was obtained from the NWC. Following harvest of material in the nursery, labelled rootstocks were retained and transferred back into a gauzehouse to allow growth of leaf material for use in subsequent laboratory-based beetle feeding and rust resistance assays (Section 4.2.7). Cuttings were maintained in cold storage at 1°C prior to planting.

Figure 4.3. The K1 and K8 mapping populations growing in the nursery at LARS.



4.2.5 Planting the field trial

Prior to planting, the trial site was prepared by marking out the area, spraying with glyphosate to remove grass weeds, harrowing and rolling. To ease planting, the ground was first softened by watering. Six cuttings per individual were planted by hand according to the field plan during 20th – 23rd March 2000 (Figure 4.4). As the quality of the cuttings was not consistent for some progeny that were low-yielding in the nursery, those with the largest diameter were preferentially planted in positions 3 and 4 of each plot (Figure 4.1), as the majority of phenotypic assessments would be based on stools growing at these positions (Section 4.2.6). Where less than six cutting were available, plot positions remained empty to prevent possible errors in future assessments and sampling. To promote establishment, the cuttings were watered immediately after planting. An electrified rabbit fence was erected to surround the entire trial site.

Figure 4.4. Planting the field trial.

4.2.6 Field-based trait assessments

Field-based phenotypic assessments were performed for several traits so that the most suitable population for future genotyping (Chapter 5) and QTL analysis (Chapter 6) could be chosen. Traits assessed in both of the K1 and K8 populations included rust resistance/susceptibility, destructive assessments of yield (fresh and dry weights) and

assessment of two important components of yield (stem height and stem diameter) (Tsarouhas *et al.*, 2002). All genotypes of the K1 and K8 progenies were included in assessments. Plots containing parental and grandparental material were also assessed. To allow identification of any across-site variation in trait values and adjustment of phenotypic values prior to QTL analysis if necessary (Section 6.2.2), trait values for the 'Jorr' individuals interspersed throughout the trial were also recorded.

Prior to all assessments relating to yield, all plots were examined for damage arising from mammalian pests such as deer, hares and rabbits. So as not to introduce erroneous trait scores into the experiment, all stools showing evidence of damage were excluded from yield-related assessments.

4.2.6.1 Field-based assessments of rust resistance

To ensure that rust infestation of the trial had progressed to a maximum level, field-based assessments of rust resistance/susceptibility were performed late in the growing season (September 2000). Plants were visually examined for the presence of uredinia and each genotype was assigned a score according to a scale developed and used in previous years' assessments by researchers at LARS (Hunter and Peacock, 2001). Scores ranged from zero, if no evidence of infection was observed, to six, where the leaf surface was almost completely covered in pustules and evidence of defoliation was observed. Scores obtained from the two stools in positions 3 and 4 of each plot were averaged to give a final value for each plot.

4.2.6.2 Destructive yield assessments

Following growth in the establishment year, all plots within the field trial were harvested during January - February 2001. Total fresh weights were recorded for stools 3 and 4 of each plot. Mean values were calculated from the two individual measurements obtained per plot.

For assessment of the dry matter yield, material from plants 3 and 4 of each plot was chipped using a standard garden chipper and transferred into individual trays for drying at 100°C for 17 h. Resulting dry material was weighed to determine the total

dry yield. These data were then used to calculate the mean percentage moisture content of each individual plot.

4.2.6.3 Stem diameter assessments

Prior to harvest of the entire trial in 2001, stem diameters were measured for all stems of stools at positions 3 and 4 for all progenies, parents, grandparents and 'Jorr' reference clones. All stem diameter measurements were recorded at an above ground stem height of 55 cm using digital calipers (Masser Caliper Model 35, Savcor Ltd.) according to the Manufacturers' instructions. For initial trait analysis described in this Chapter, the mean of the maximum stem diameter for each of the two stools measured per plot was used.

4.2.6.4 Stem height assessments

Two independent assessments of stem height were performed. The first was based on the single stems of each individual K1 and K8 genotype following the initial season of growth in the nursery. For the second assessment, based on the populations growing in the field trial, the heights of all stems per stool were recorded for plants 3 and 4 of each plot assessed. This latter assessment was performed immediately prior to the destructive yield assessments. To obtain a final trait value, the mean value of the maximum stem height of plants 3 and 4 was calculated.

4.2.7 Laboratory-based trait assessments

In addition to the field-based trait assessments outlined in Section 4.2.6, two laboratory-based experiments were performed. The first was a rust resistance assay based on leaf disc inoculation techniques developed at LARS (Pei *et al.*, 1997). The second aimed to identify any differences among individuals of the mapping population in their attractiveness to the leaf-feeding invertebrate pest, the blue willow beetle (*Phratora vulgatissima*).

4.2.7.1 Laboratory-based assessment of rust resistance

Laboratory-based assessments of rust resistance were performed essentially as described by Pei *et al.* (1996). In brief, leaf discs (1.1 cm diameter) were prepared from 100 randomly selected individuals from the K1 and K8 populations and 100 individuals from the K3 population, to include the 66 original progeny used to construct the K3 linkage map and 34 randomly selected additional progeny (Chapter 3) (Table 4.3). For K1 and K8, leaf material was obtained from regenerated rootstocks growing in a gauzehouse. For K3 individuals 1 – 66, five cuttings per individual were grown in six-inch diameter pots. Additional K3 material (individual 67 onward) was obtained from regenerated rootstocks.

Table 4.3. Individuals selected from the K1, K8 and K3 mapping populations for inclusion in rust inoculation tests.

Population	Individuals included
K1	2, 7, 17, 17, 25, 32, 40, 44, 48, 53, 62, 86, 92, 94, 109, 113, 115, 117, 131, 144, 151, 164 ,165, 177, 179, 185, 218, 241, 244, 256, 259, 262, 268, 276, 297, 299, 308, 387, 389, 402, 426, 440, 447, 454, 471, 476, 496, 513, 519, 527, 530, 533, 542, 548, 557, 562, 567, 568, 580, 600, 605, 619, 631, 665, 668, 686, 694, 722, 724, 732, 750, 758, 759, 770, 778, 780, 782, 802, 815, 821, 829, 834, 841, 844, 846, 849, 852, 865, 868, 877, 889, 892, 893, 899, 903, 907, 908, 925, 928, 933, 936
K8	5, 19, 30, 35, 37, 43, 46, 53, 71, 78, 81, 82, 87, 114, 125, 134, 137, 141, 160, 167, 172, 177, 178, 211, 215, 219, 223, 229, 232, 235, 247, 248, 250, 266, 280, 282, 331, 333, 334, 349, 357, 360, 365, 400, 412, 428, 438, 439, 443, 446, 453, 470, 473, 488, 495, 496, 502, 507, 511, 520, 530, 542, 550, 556, 574, 588, 591, 596, 599, 619, 626, 651, 652, 655, 676, 706, 711, 718, 756, 766, 769, 776, 785, 810, 814, 830, 836, 845, 850, 856, 860, 878, 889, 896, 881, 904, 918, 920, 928, 929
K3	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 68, 76, 79, 82, 91, 93, 99, 100, 101, 103, 104, 108, 111, 112, 113, 118, 126, 133, 137, 150, 153, 154, 163, 167, 170, 175, 178, 184, 188, 190, 195, 196, 202, 205

Leaves five - ten (where one is the first unfurled leaf, defined when the furled edge is less than one-third of the entire leaf edge) were used. Leaf discs (1.1 mm^2) were placed on tap water-soaked paper bridges, abaxial side uppermost, in 10 x 10 cm repli-plates (Sterilin). Prior to inoculation, isolates of LET1 (isolate VMP891-1R; kindly provided by M. Pei) and LET5 (isolate STP895-11; M. Pei) were multiplied by re-inoculation of detached willow leaves floating on tap water in Petri-dishes at 18–20 °C in an illuminated growth chamber. Spore suspensions of 1 – 2 x 10^5 viable spores per

ml were prepared in tap water containing 0.004% Tween 20 (Sigma) and inoculated onto the leaf discs using a Humbrol air brush (Humbrol Ltd, UK). Following inoculation, the leaf discs were incubated at 16°C with 16 h/d illumination at an intensity of 80 $\mu E\ m^{-2}\ s^{-1}$ for 13 days. Leaf discs were then digitally photographed to provide a permanent record of the results.

To obtain a rust resistance/susceptibility score for each individual, the number of pustules on each of the five replicate leaf discs was recorded and a mean value calculated per genotype.

4.2.7.2 Laboratory-based beetle feeding experiments

The beetle feeding experiment was based on a protocol previously developed at LARS (Wiltshire *et al.*, 1997). This approach aims to identify feeding preferences of willow beetles for a particular willow genotype when presented with a selection of different genotypes on which to feed.

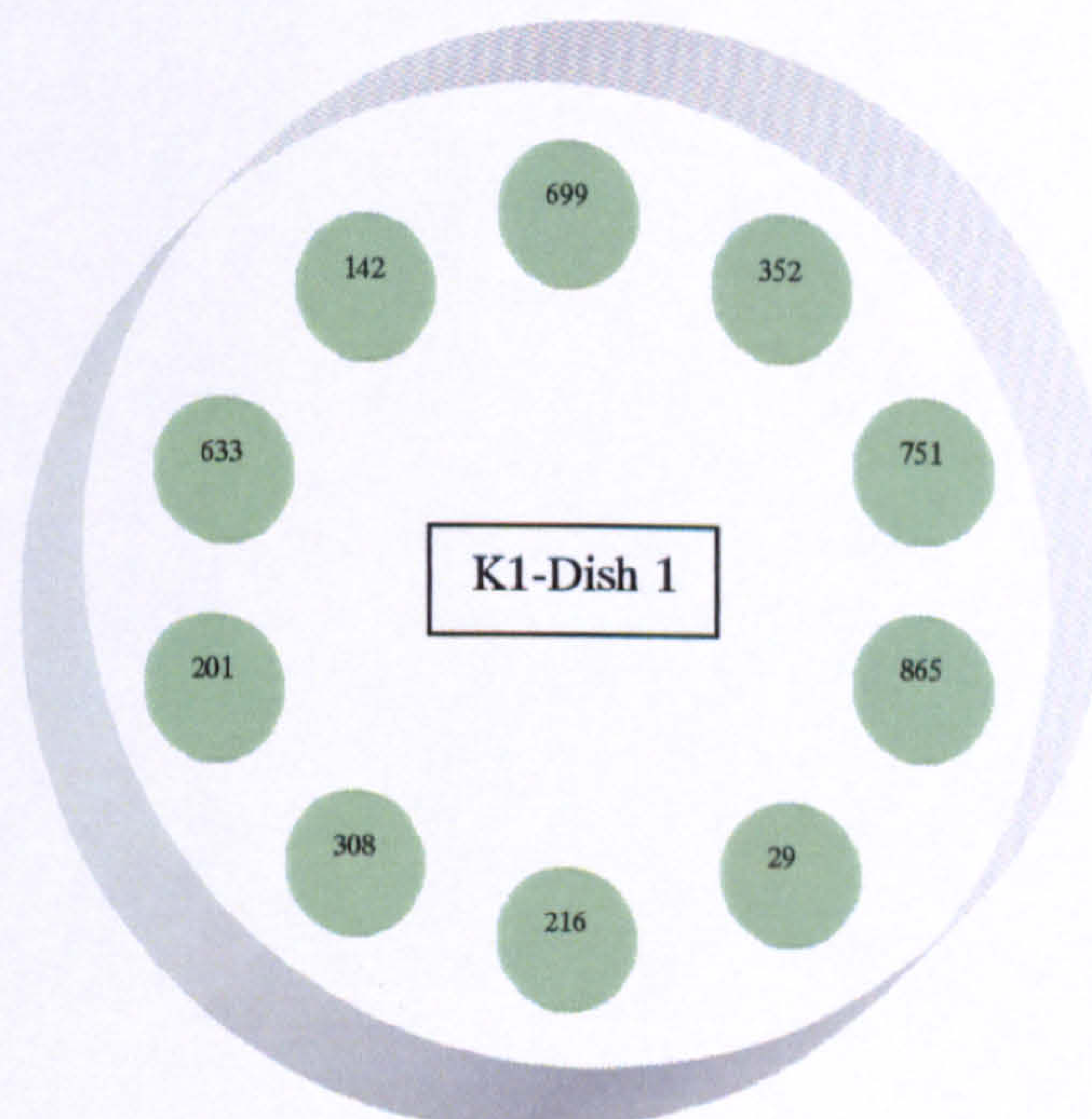
For the K1 and K8 progenies, 89 individuals were randomly selected (Table 4.4) for testing. From the K3 population, 89 individuals were also selected to included progeny numbers 1 – 66 and 23 additional individuals randomly selected from the numbers 67 – 209. Parental material for each of the populations was also included in each respective experiment. The experimental design was based on an incomplete Latin square model to ensure that a leaf disc from a given tested individual never occurred in a Petri-dish with a disc from any of the other tested individuals more than once. This approach aimed to eliminate the possible effects arising from the occurrence of competing choices within the same Petri dish. The layout of leaf discs within dishes is provided in Appendix I.

Table 4.4. K1, K8 and K3 individuals included in laboratory-based beetle feeding assays.

Population	Individuals included
K1	25, 29, 33, 52, 74, 78, 81, 82, 89, 93, 115, 116,133, 142, 159, 164, 168, 171, 177, 199, 201, 205, 216, 225, 229, 235, 236, 244, 252, 256, 266, 299, 300, 305, 308, 309, 320, 324, 328, 352, 353, 374, 410, 420, 424, 425, 433, 451, 463, 475, 488, 502, 518, 538, 554, 555, 603, 607, 614, 615, 633, 652, 654, 658, 664, 669, 680, 683, 686, 699, 702, 715, 743, 751, 776, 786, 788, 806, 820, 819, 840, 860, 865, 868, 874, 877, 919, 929, 939
K8	10, 19, 26, 46, 50, 71, 77, 84, 94, 106, 127, 132, 166, 176, 180, 183, 189, 192, 197, 230, 237, 241, 256, 287, 292, 297, 296, 309, 315, 319, 323, 326, 347, 350, 359, 379, 381, 386, 414, 420, 421, 422, 434, 436, 438, 440, 460, 470, 503, 507, 517, 520, 542, 547, 548, 555, 577, 581, 583, 597, 608, 614, 628, 648, 695, 728, 732, 759, 763, 767, 772, 774, 781, 786, 805, 817, 825, 841, 866, 868, 880, 905, 909, 912, 915, 920, 921, 931, 939
K3	1 - 66, 74, 75, 88, 96, 98, 100, 101, 108, 114, 118, 128, 131, 132, 136, 142, 146, 158, 164, 175, 180, 188, 203, 204

Due to the large scale of the experiment it was necessary to assay each population at a different time. On the day of each experiment, 910 adult *P. vulgatissima* beetles were collected from willows growing at LARS and starved for 6 h prior to release. Ten leaf discs per plant (1.2 cm diameter) were cut from the selected individual regenerated rootstocks growing in the glasshouse and pinned, using fine entomological pins (Watkins Ltd, Doncaster) at even intervals, axial side uppermost, to labelled filter papers (Whatman) in 22 cm diameter Petri-dishes (Sterilin) according to the model design (Appendix I) as illustrated in Figure 4.5.

Figure 4.5. Schematic representation of leaf disc layout for Petri-dish 1, population K1.



To avoid desiccation of leaf discs during the experiment, 5 ml of tap water was added to each dish. Ten beetles were released into each dish and allowed to feed for 16 h. Leaf discs were then removed and digitally photographed to provide a permanent record of the results. The amount of leaf area remaining for each leaf disc was then recorded, using an Optomax V Image Analyser (Synoptics, Cambridge), and the mean for each individual was calculated from the ten replicates per genotype.

To allow for a more informed choice of population for subsequent use in this study, resulting assessment values for the K1 and K8 experiments were analysed using analysis of variance (ANOVA) in Genstat (Section 2.7.3) to test whether resulting differences in area of leaf remaining between individuals were significant in each population, i.e. observed differences were not due to experimental inconsistencies only.

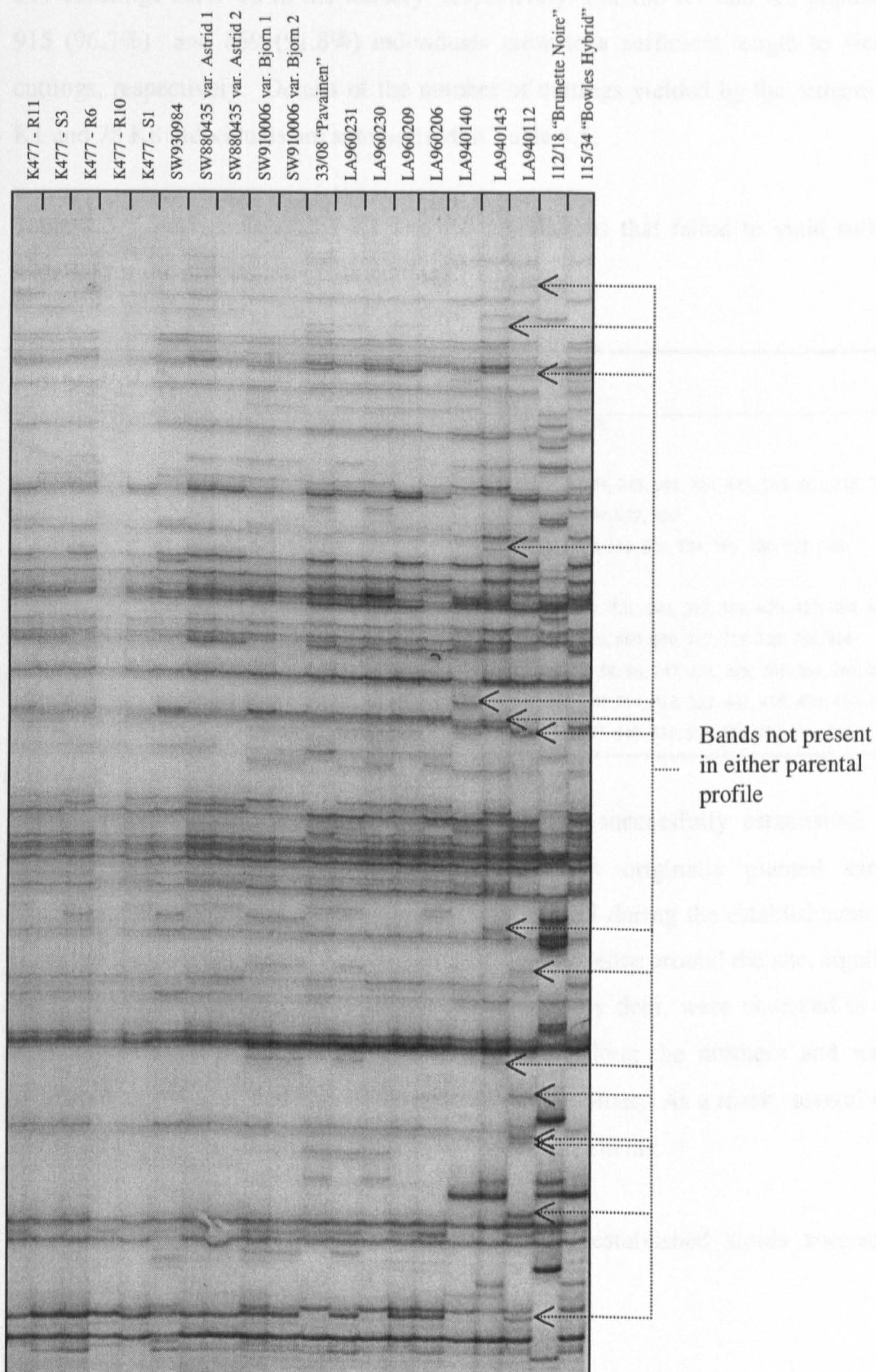
4.3 Results

4.3.1 AFLP analysis for authentification of parents used in candidate pedigrees

AFLPs were used to check the identification of the parents of seven candidate crosses against their known parents and grandparents in the NWC. Both primer combinations used for AFLP analysis successfully generated high quality profiles. The two independent profiles generated using different DNA extracts for the varieties Björn and Astrid were identical for both samples, indicative of the robustness of the experiment. In profiles generated with both primer combinations, several clear bands were detected in the full-sib individuals LA940140, LA940143 and LA940112 that were not present in either of the parental profiles (112/18 "Brunette Noire" and 115/34 "Bowles Hybrid") (Figure 4.6). Furthermore, several bands detected in 112/18 "Brunette Noire" were not detected in any of the three progeny analysed (Figure 4.6). With the exception of a single band, all bands detected for the second expected parent (115/34 "Bowles Hybrid") were also detected in the profile of at least one of the three offspring.

In contrast, all AFLP bands in individuals of the K477 population were present in either of the parents SW930984 or SW880435 var. Astrid. Similarly, all bands detected in LA960230 and LA960231 (cross 581; Table 4.1) were also observed in their parents (SW910006 var. Björn and 33/08 "Pavainen").

Figure 4.6. An AFLP autoradiograph showing profiles generated for all candidate parents and grandparents of the mapping population. Pedigree information is provided in Table 4.1.



4.3.2 Establishment of the field trial

Of the 950 seeds randomly selected for each of the K1 and K8 populations, 946 and 947 seedlings survived in the nursery, respectively. For the K1 and K8 populations, 915 (96.7%) and 869 (91.8%) individuals grew to a sufficient length to yield six cuttings, respectively. Details of the number of cuttings yielded by the remaining 31 K1 and 78 K8 individuals are summarised in Table 4.5.

Table 4.5. Individuals of the K1 and K8 populations that failed to yield sufficient material for the production of six cuttings.

Number of cuttings	Population	
	K1	K8
1	118, 275, 827	263, 758
2	163, 818	162, 238, 245, 291, 324, 425, 563, 586, 716, 733, 791, 890, 899, 927, 930
3	83, 265, 255, 643, 762, 858	85, 269, 332, 538, 724, 755, 780, 921, 940
4	237, 298, 332, 359, 479, 581, 764, 905, 906	14, 158, 221, 243, 273, 382, 419, 427, 454, 480, 567, 595, 615, 689, 699, 717, 719, 725, 768, 914
5	15, 329, 427, 545, 616, 787, 771, 826, 875, 898, 938	11, 28, 58, 86, 147, 181, 225, 261, 264, 265, 268, 283, 300, 310, 314, 318, 322, 401, 415, 430, 435, 441, 487, 497, 507, 518, 531, 554, 578, 754, 760, 851

For all populations, a high proportion of cuttings successfully established. For example, 5322 (~97%) of the 5513 K8 cuttings originally planted survived establishment. In general, both populations grew well during the establishment year. However, despite the presence of an electrified rabbit fence around the site, significant levels of mammalian damage, attributed to grazing by deer, were observed in some areas of the field trial. This was most prevalent along the northern and western perimeters of the trial, and also in the south-western corner. As a result, several stools were excluded from subsequent yield-based trait assessments.

Following harvest after the first year of growth, established stools successfully regenerated as shown in Figure 4.7.

Figure 4.7. The field trial at LARS in the second year of growth.

4.3.3 Trait assessments

4.3.3.1 Field-based assessments

In the K1, K8 and K3 mapping populations phenotypic variation was observed for all field-based traits assessed. The distributions of raw, unadjusted scores for each trait assessed are shown in Figure 4.8. For all traits assessed, with the exception of mean number of stems per stool, a wider distribution of phenotypic values was observed for each trait in the K1 and K8 populations in comparison to K3. Furthermore, again with the exception of mean number of shoots per stool, greater variation for all traits assessed was observed in K1 than in K8. When comparing trait distributions between populations K1 and K8, phenotypic values were, in general, more evenly distributed across the range of observed values in K8 than in K1, i.e. for the rust assessment, rust scores between 0 and 1 were recorded for 66.5% of K1 individuals assessed. However, for K8, 39.5% of rust scores were in this category. A similar trend was observed for all other traits assessed.

For all traits, variation in the values recorded for the reference Jorr plots within the K1 and K8 populations, was observed (Table 4.6). However, for each trait, the level of variation observed in such plots was less than that observed within the K1 and K8 progenies. Furthermore, with the exception of fresh weight, greater variation was observed for all traits in Jorr plots interspersed throughout the K1 population than for those within K8.

Summarised trait data for the K1 and K8 mapping populations, their respective putative parents and grandparents and the reference Jorr plots within each population are shown in Table 4.6.

Figure 4.8. Field- and nursery-based trait value distribution plots.

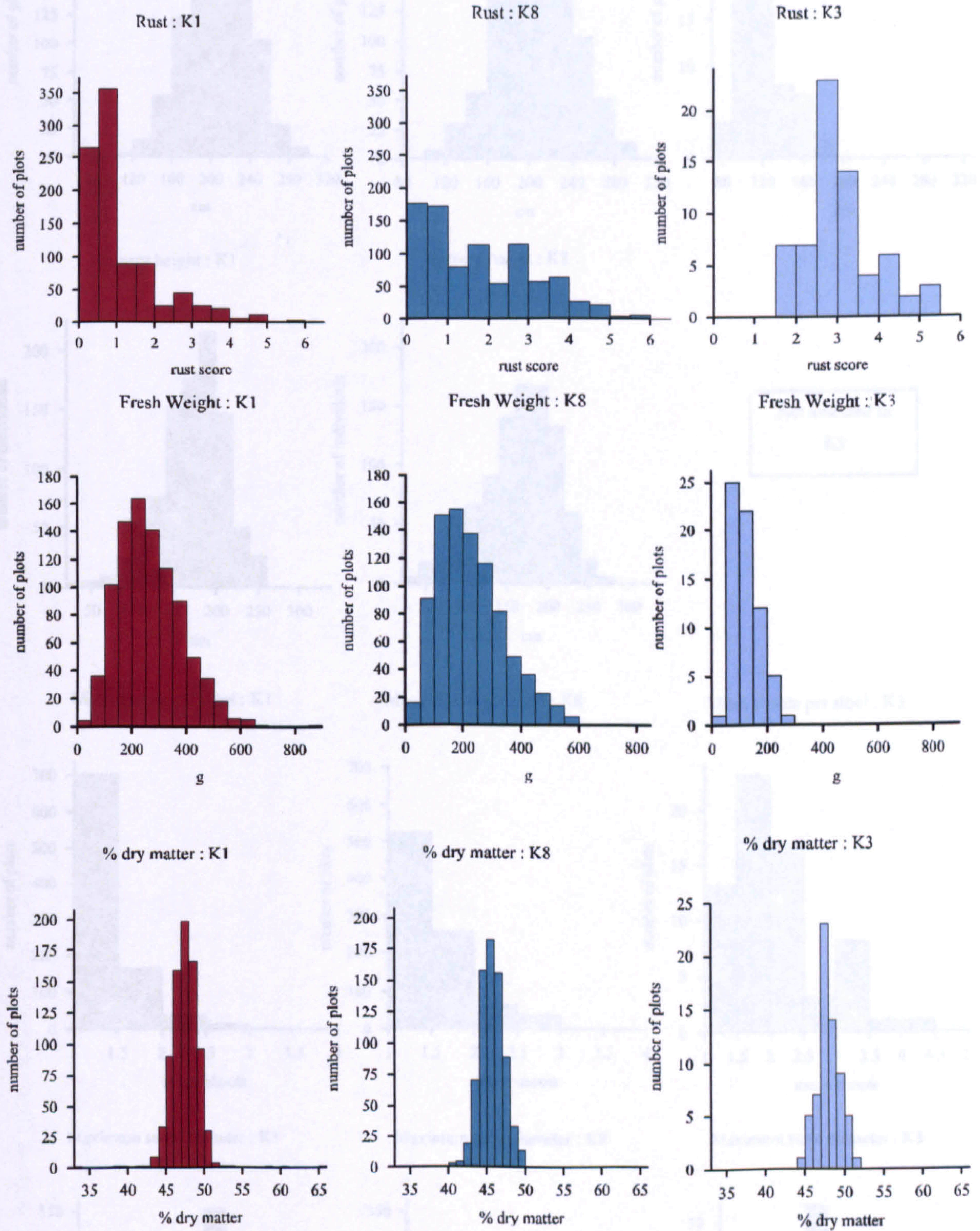


Figure 4.8 continued.

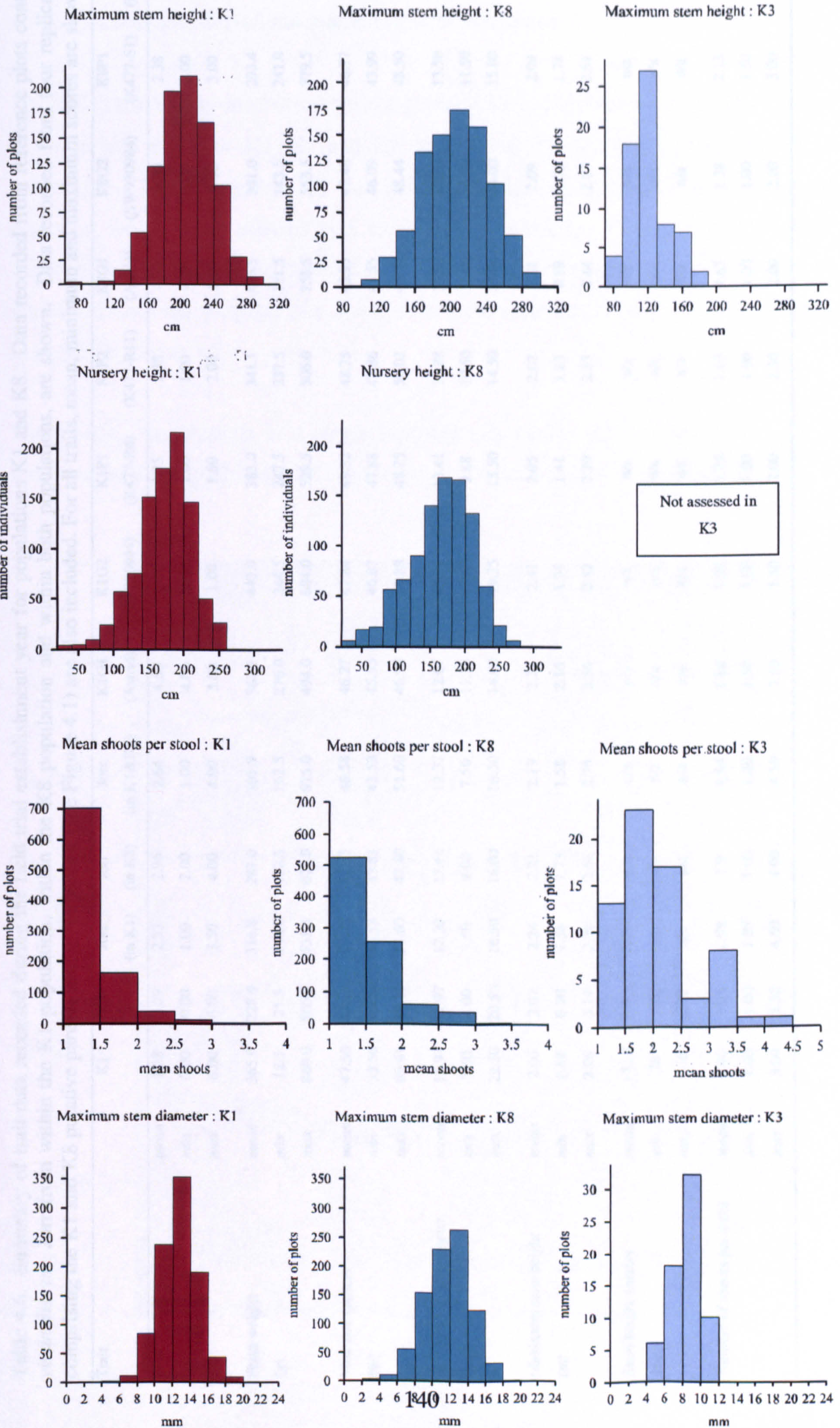


Table 4.6. Summary of trait data recorded during the field trial establishment year for populations K1 and K8. Data recorded from reference plots containing *S. viminalis* var. Jorr from within the K1 population, within the K8 population and within both populations, are shown. Data recorded from four replicated plots comprising the K1 and K8 putative parents and grandparents (see Figure 4.1) are also included. For all traits, mean, minimum and maximum scores are shown.

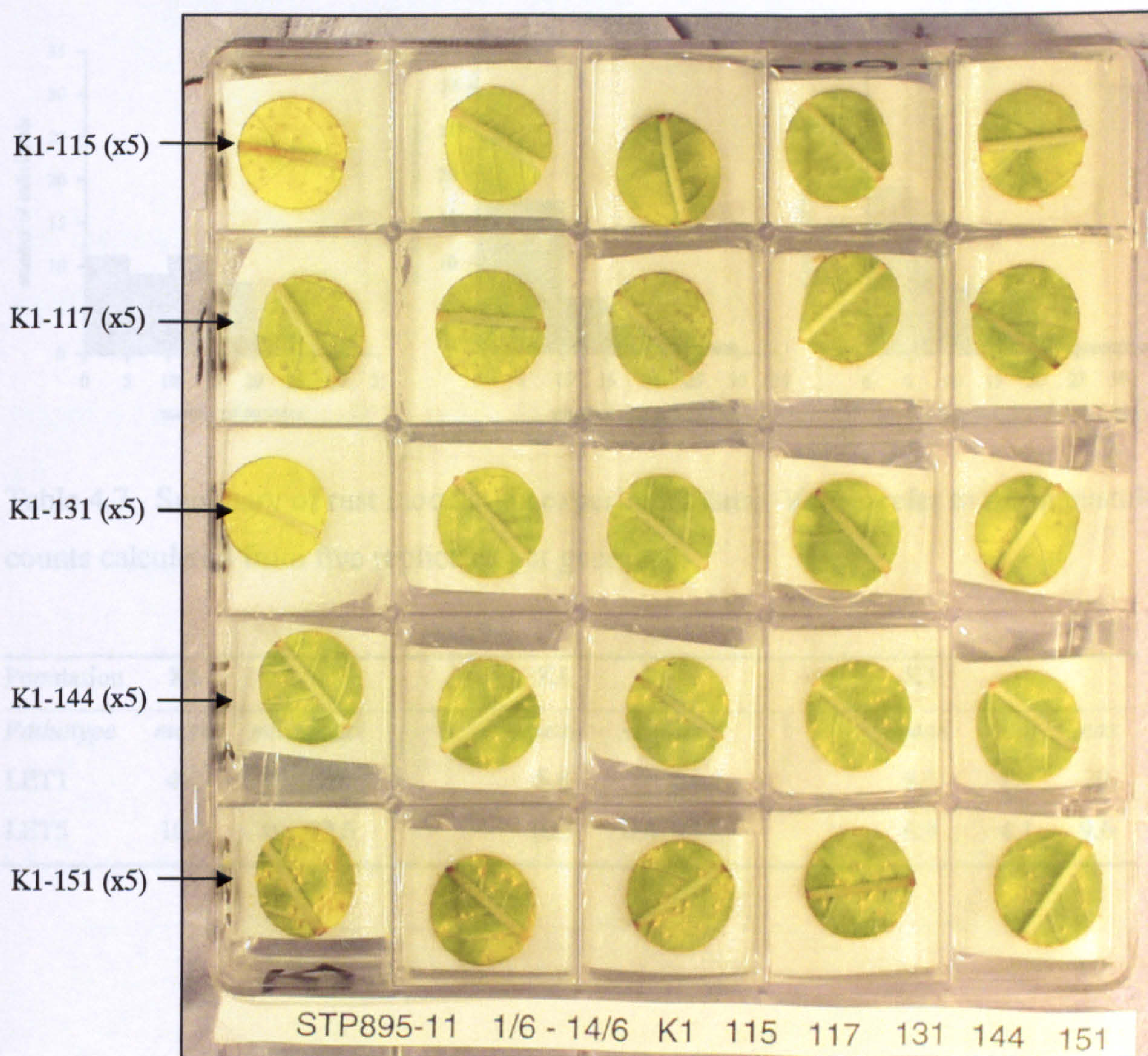
Trait		K1	K8	Jorr	Jorr	Jorr	K1G1	K1G2	K1P1	K1P2	K8G1	K8G2	K8P1	K8P2
				(in K1)	(in K8)	(in K1&K8)	(Astrid)	(SW930984)	(K477-R6)	(K477-R11)	(Astrid)	(SW930984)	(K477-S1)	(K477-R11)
Rust	*mean*	1.28	2.09	2.33	2.96	2.64	4.50	0.50	1.25	1.38	4.50	0.00	2.38	0.38
	min	0.00	0.00	1.00	2.00	1.00	4.00	0.00	1.00	1.00	4.00	0.00	2.00	0.00
	max	6.00	6.00	3.50	4.00	4.00	5.00	1.00	1.50	2.00	5.00	0.00	3.00	1.00
Fresh weight	*mean*	265.9	228.9	316.8	287.0	301.9	367.1	443.9	383.0	341.1	446.5	301.0	293.4	324.5
(g)	*min*	12.5	25.5	140.0	102.5	102.5	276.0	244.5	247.5	237.5	171.5	143.5	241.0	257.0
	max	860.0	805.5	531.5	675.0	675.0	494.0	684.0	529.5	506.0	758.5	553.5	379.5	393.5
Moisture content	*mean*	47.50	45.65	45.77	45.63	46.58	46.27	47.84	48.32	48.75	45.95	47.48	46.77	48.87
(%)	*min*	33.56	40.24	43.53	43.81	43.53	45.55	46.87	47.88	47.86	45.55	46.09	43.99	48.11
	max	65.61	53.11	51.60	47.40	51.60	46.98	48.89	48.75	50.02	46.77	48.44	48.50	49.80
Maximum stem diameter	*mean*	12.97	11.97	12.30	12.44	12.37	12.67	15.31	13.41	12.88	13.50	12.33	13.50	14.00
(mm)	*min*	3.50	4.00	7.50	8.00	7.50	11.25	13.50	9.88	10.50	11.00	10.50	11.50	12.50
	max	22.50	20.50	16.50	16.00	16.50	14.00	19.25	15.50	14.50	17.00	14.00	15.00	16.00
Maximum stem height	*mean*	2.06	2.07	2.24	2.21	2.19	2.27	2.41	2.05	2.02	1.81	2.09	2.09	1.96
(m)	*min*	0.68	0.90	1.58	1.73	1.58	2.16	1.98	1.41	1.83	0.89	1.73	1.78	1.85
	max	2.96	3.14	2.76	2.68	2.76	2.36	2.82	2.79	2.33	2.44	2.83	2.51	2.14
Stem height nursery	*mean*	172.5	165.1	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
(cm)	*min*	26	26	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	max	336	272	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Number of shoots per stool	*mean*	1.45	1.63	1.98	1.9	1.94	1.88	1.38	1.75	1.63	1.63	1.38	2.13	2
	min	1.00	1.00	1.25	1.00	1.00	1.50	1.00	1.00	1.00	1.00	1.00	1.50	1.50
	max	3.00	3.50	4.50	4.00	4.50	2.50	1.50	2.00	2.50	2.00	2.00	3.00	2.50

4.3.4 Laboratory-based trait assessments

4.3.4.1 Laboratory-based assessment of rust resistance

Following the 13 day incubation period, pustules were observed on several of the leaf discs derived from the K1, K8 and K3 populations with both the LET1 and LET5 pathotypes. For populations K1 and K8, several leaf discs showed no evidence of infection, while others showed high levels of susceptibility. However, pustules were observed on all individuals from the K3 population, although the level of infection varied between the progeny. An example of results obtained from a single repli-plate is shown in Figure 4.9.

Figure 4.9. An example of a single repli-plate showing variation in the number of observed uredinia (pathotype LET5) for five individuals of the K1 population.



The distribution of mean pustule counts for each population tested and for each inoculum used is shown in Figure 4.10.

Figure 4.10. Distribution plots for mean pustule count data for the K1, K8 and K3 populations, following inoculation with two rust pathotypes (LET1 and LET5).

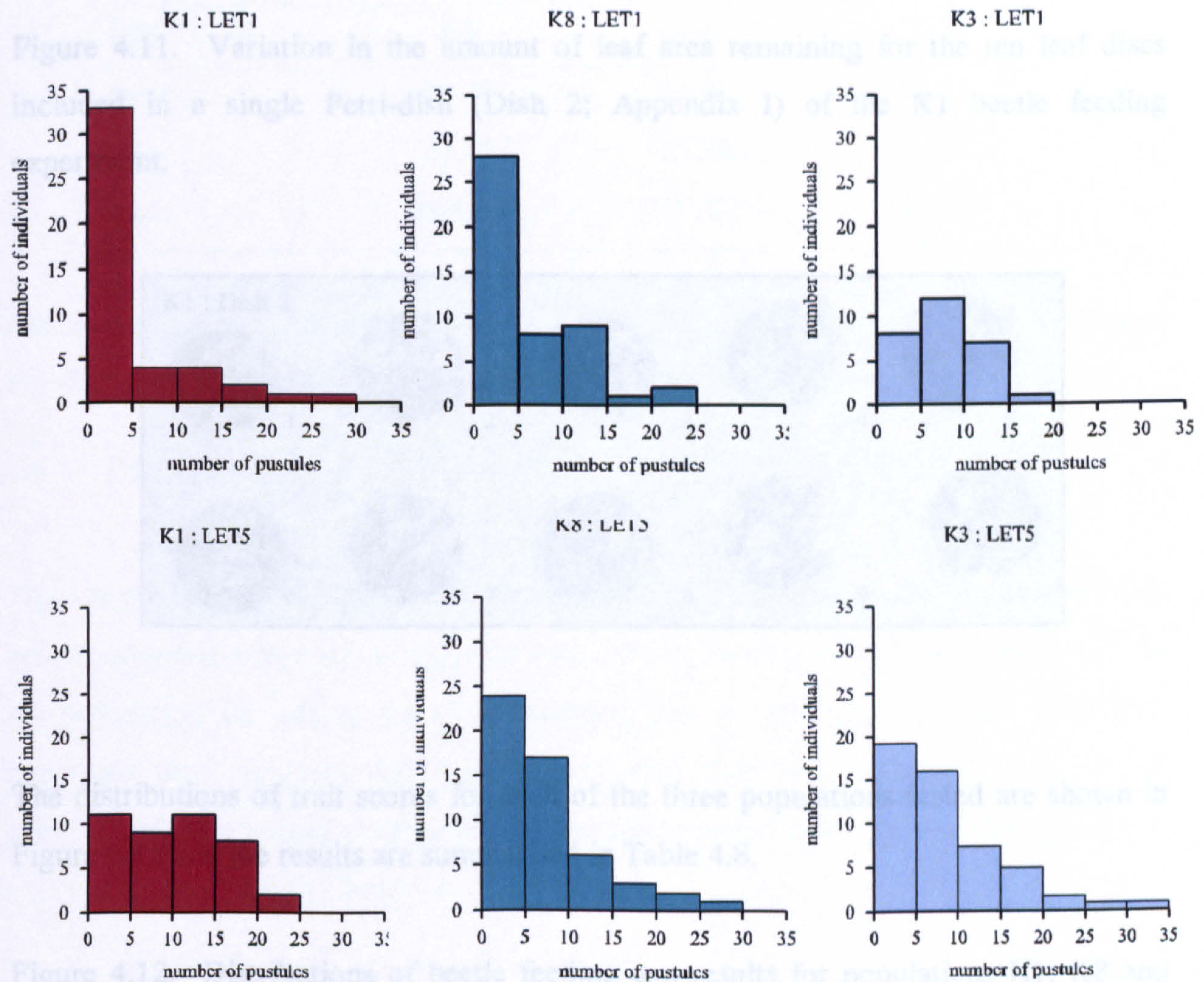


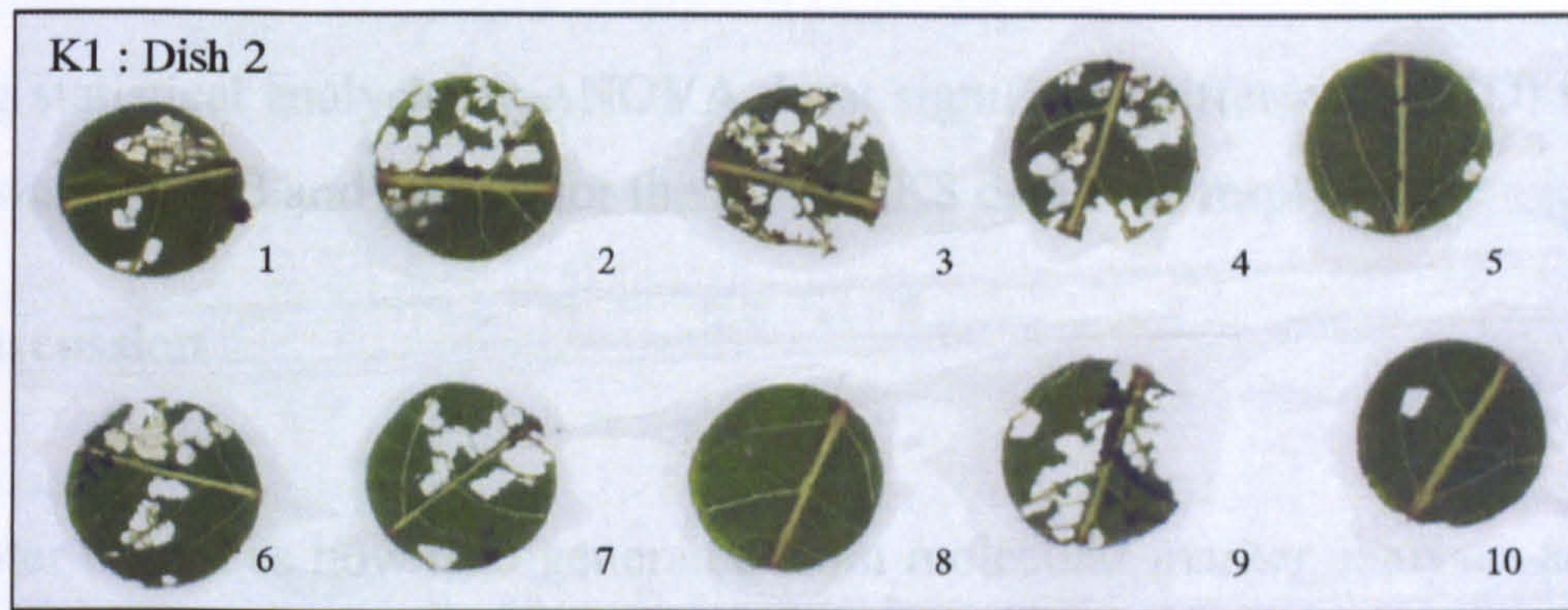
Table 4.7. Summary of rust inoculation experiment data. Values refer to mean pustule counts calculated from five replicates per genotype.

Population	K1			K8			K3		
Pathotype	*mean*	*min*	*max*	*mean*	*min*	*max*	*mean*	*min*	*max*
LET1	4.2	0	29	5.8	0	24.2	8.3	1.4	20
LET5	10.3	0	22.6	10.3	0.6	22.6	6.9	4.1	9.6

4.3.4.2 Laboratory-based beetle feeding experiments

For each experiment, variation in the amount of leaf area remaining was observed between individual leaf discs of the K1, K8 and K3 progenies. An example of the results from a single dish is shown in Figure 4.11.

Figure 4.11. Variation in the amount of leaf area remaining for the ten leaf discs included in a single Petri-dish (Dish 2; Appendix I) of the K1 beetle feeding experiment.



The distributions of trait scores for each of the three populations tested are shown in Figure 4.12 and the results are summarised in Table 4.8.

Figure 4.12. Distributions of beetle feeding test results for populations K1, K8 and K3.

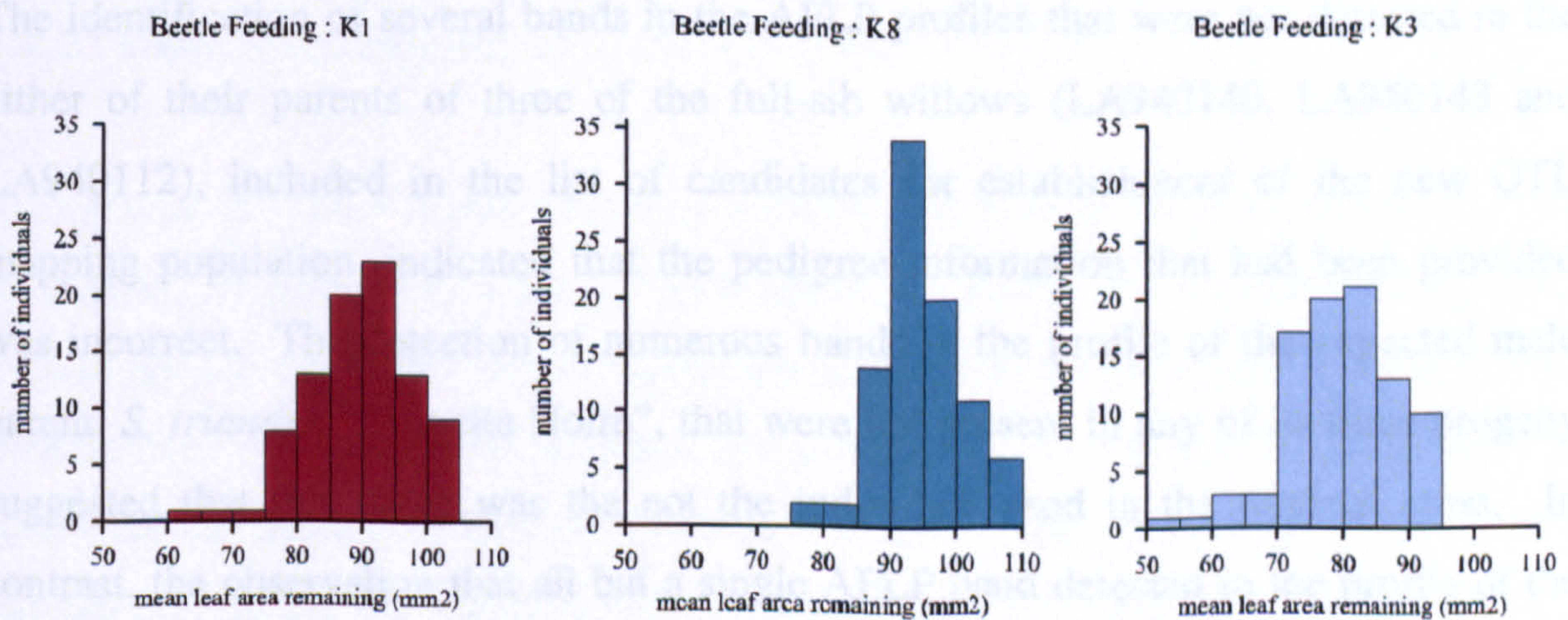


Table 4.8. Summary of results from beetle feeding assays for populations K1, K8 and K3. Values indicate the mean amount of leaf area remaining (mm^2) calculated from the ten replicates per genotype.

Population	K1	K8	K3
mean	89.29	94.49	79.92
min	64.48	75.50	51.10
max	103.09	107.87	100.75

Following statistical analysis by ANOVA, least significant difference (LSD) values at $P = 0.05$ were 13.353 and 10.213 for the K1 and K8 data sets, respectively.

4.4 Discussion

The Chapter describes how data generated from molecular marker analyses and field- and laboratory-based trait assessments were used in making an informed choice of full-sib population for use in the QTL mapping part of this present study. All data generated from analyses and assessments were, therefore, addressed only in this context. More detailed examination of the fundamental information afforded by the trait data is provided in Chapter 6.

4.4.1 AFLP analysis of candidate pedigrees

The identification of several bands in the AFLP profiles that were not detected in the either of their parents of three of the full-sib willows (LA940140, LA940143 and LA940112), included in the list of candidates for establishment of the new QTL mapping population, indicated that the pedigree information that had been provided was incorrect. The detection of numerous bands in the profile of the expected male parent, *S. triandra* "Brunette Noire", that were not present in any of its three progeny suggested that this clone was the not the individual used in the original cross. In contrast, the observation that all but a single AFLP band detected in the profile of the proposed female parent were also detected in at least one of the three progeny tested, may suggest that this clone is a true parent. These results led to the exclusion of

crosses 524, 529, and 553 from those that may have eventually been used as the mapping population.

All AFLP bands detected in the K477 population individuals (R11, S3, R6, R10 and S1) were also detected in the parental (SW930984 and SW880435 var. Astrid) and grandparental (SW910006 var. Björn and SW880435 var. Astrid) profiles. This was indicative of the robustness of this pedigree. This was also found with the full-sib willows LA960230 and LA960231 used to make cross 581 (Table 4.1).

The value of molecular markers in pedigree testing was clearly demonstrated in this phase of the study, with the results obtained from AFLP analysis playing a key role in the final selection of candidate mapping populations.

4.4.2 Field-based trait assessments

For all field traits assessed, inter-population comparisons of trait values were performed based on the raw data recorded in the field and did not, therefore, take account of any positional variation that may have been present within the trial site. While the adjustment of recorded trait values according to any spatial heterogeneity observed across the Jorr reference plots may have permitted a more informed choice of the mapping population, limitations regarding both time and statistical support prevented such an approach being employed. However, where applicable, trait data generated in this phase of the project was later adjusted for spatial variation prior to use in QTL analysis (Section 6.2.2).

4.4.2.1 Field-based assessment of rust resistance

Rust scores in both K1 and K8 progenies ranged across all possible scores, but the replicated Jorr plots yielded a narrower range of rust scores in comparison (Table 4.6). This suggested that genes conferring resistance to rust infection are segregating in both populations. Although no detailed analysis aimed at the identification of any positional trends relating to rust score was attempted during this phase of the project, the fact that all the replicated Jorr plots showed signs of rust infection suggested that the entire field trial site had been subject to disease. Moreover, a high degree of consistency in rust scores for Jorr plots was observed, with 87% of assigned scores

between two and three. Further evidence that the variation in rust scores was not wholly due to positional effects was provided by the observation that several plot positions containing highly susceptible mapping population individuals were located adjacent to highly resistant plots, e.g. a mean rust score of five was recorded for K1 individual 53. The mean rust scores for the two adjacent plots, 52 and 54, were both zero.

The mean rust score for the K8 population as a whole (2.09) was greater than the mean rust score of the K1 population (1.28). This may reflect differences in the genic contribution of rust resistance genes and alleles between the two crosses. However, the mean rust score calculated from all of the Jorr reference plots within the K1 population (2.33) was also slightly less than the corresponding mean score obtained for the K8 population (2.64), although the statistical significance of this difference was not tested. If found to be significant, this result may be indicative of spatial inconsistencies relating to disease pressure. Alternatively, this may be a consequence of more favourable conditions for pathogen establishment resulting from the overall increased susceptibility of the K8 population in comparison to K1.

Relative mean rust scores obtained from the parents and grandparents of both the K1 and K8 populations were in general agreement with data previously generated in Swedish assessments (Tables 4.2 and 4.6), with SW880435 var. Astrid consistently being most susceptible and SW930984 always the most resistant. As expected, K477-R6 and K477-R11 were more resistant to rust than K477-S1. This may indicate that disease pressures in both the UK and Sweden were of a similar nature at the time of each respective assessment, i.e. similar rust pathogens, with comparable pathogenicities were present.

The assessment results suggest that both the K1 and K8 progenies are suitable for QTL analysis aimed at identifying genomic regions underlying rust resistance. However, the wider distribution of rust scores attained for K8 progeny may be a consequence of segregation of additional alleles that do not segregate in family K1, thus, K8 may be a more informative population for mapping QTL involved in resistance. In contrast to populations K1 and K8, the fact that none of the K3 progeny assessed demonstrated complete resistance (Figure Table 4.7) indicates that this population would be an inferior choice, in comparison to K1 or K8, for mapping QTL involved this trait.

4.4.2.2 Assessments relating to yield

Although variation recorded for yield assessments in the Jorr reference plots within the K1 and K8 population was, with the exception of mean number of stems per stool, consistently less than that observed within the mapping populations, the wide variability in trait scores recorded for these clonal replicates suggested that environmental factors were contributing to the phenotypic variation observed across the K1 and K8 progenies. While genetic factors may also have influenced phenotype, without adjustment of trait values according to any spatial heterogeneity that may have be present, differentiation of genetic effects from environmental influences was difficult. In turn, accurate inter-population comparisons of yield-based traits, in terms of underlying genetics, were confounded. However, some potentially informative general trends were observed. First, the more even distribution of yield-related trait values in K8 compared to K1 may suggest that a greater number of QTL affecting these traits are segregating in population K8. Additional support for this hypothesis was provided by the observation that, apart for fresh weight, greater variation in all yield-related trait values were recorded for the Jorr reference plots located within the K1 population than for those within K8. Had greater variation been observed between Jorr plots in K8 compared to those in K1, then the wider distribution of trait values in K8 could have occurred purely as a consequence of more variable environmental influences in this area of the trial.

Within the limitations of obtaining information on yield, these results also suggested that the K8 population would be a better choice for QTL mapping, because of the more even distribution of variation found in this population compared with K1.

4.4.3 Laboratory-based assessments

Laboratory-based assessments were performed to provide a more accurate reflection of differences in the segregation of traits between the various progenies. In all cases, more standard conditions in the laboratory in comparison to the field should provide trait values that are influenced by environment to a lesser extent than related assessments performed in the field.

4.4.3.1 Laboratory-based assessment of rust resistance

Laboratory-based assessments of rust resistance were performed for two main reasons. First, as a standard inoculum could be applied in the laboratory, this approach provided a more informative system for assessing the interaction of each individual genotype with the rust pathogen. In contrast, results generated in the field-based assessment are subject to more environmental variation, particularly in infection pressures in different areas of the trial. By eliminating this variation in the laboratory, a potentially more accurate measurement of rust resistance was achieved. Second, it is known that rust populations may be highly heterogeneous within willow plantations (Samils *et al.*, 2001), with several different pathotypes being present at any given time. Therefore, field assessments of rust resistance may be simultaneously based on resistance/susceptibility to several different pathotypes. The inoculation technique permitted the examination of pathotype-specific interactions. Using two different prevalent pathotypes as inocula (LET1 and LET5) in separate experiments, potentially greater dissection of the rust resistant phenotype may be achieved, i.e. if different loci were involved in resistance to alternative pathotypes, then the detection of pathotype-specific QTL may be possible.

The results of the inoculation experiments performed in this study indicated that genes underlying resistance to both pathotypes are segregating in all three populations. However, in agreement with the field-based assessment of rust resistance, no progeny of the K3 population demonstrated complete resistance to disease, suggesting that this population may not be the most suitable for use in subsequent QTL studies. In contrast, both K1 and K8 populations showed a wide distribution of different phenotypes, indicative of the potential usefulness of these populations for mapping loci underlying resistance. When comparing K1 and K8, the greater distribution of phenotypic values recorded for K8 than K1, which was also in agreement with data recorded in the field, suggested that population K8 may be more informative as a basis for subsequent QTL mapping of this trait.

4.4.3.2 Laboratory-based beetle feeding experiments

Assessments of beetle feeding preferences were performed in the laboratory, in preference to the field, as field assessments are based on the overall levels of leaf

damage observed. Hence, resulting data can be confounded by herbivory from other leaf-feeding insect pests. Furthermore, as several willow beetle species may be present within a plantation at a given time, and different species may exhibit contrasting feeding preferences (Section 1.2.3.2.1), it was important to examine this trait based on experiments that permitted the scrutiny of feeding preferences of single beetle species. The blue willow beetle (*P. vulgatissima*) was used in these experiments since it is currently the most important invertebrate pest of biomass willows. Moreover, its relative abundance on willows growing at LARS permitted the collection of sufficient numbers of specimens for use in the experiments.

When choosing the K1 and K8 progenies as the two candidate mapping populations for use in this study, it was predicted that there would be little segregation of palatability for a single willow beetle species within these populations, owing to the large contribution of *S. viminalis* to the pedigree. As discussed in Section 1.2.3.2.1, this species is considered to be, on the whole, susceptible to damage by *P. vulgatissima*. However, also included in the pedigree of these populations is the species *S. schwerinii*, for which there remains a paucity of information in the literature regarding the relative susceptibility or resistance of this species to herbivory in relation to each of the major willow beetle species. Based on the data available for *S. viminalis*, the results of the beetle feeding tests performed here were, in general, as expected, with all individuals from all three populations showing a degree of susceptibility. As a consequence, more in-depth statistical analysis was employed, aimed at determining the robustness of the experiment and establishing whether the between-progeny variability observed was statistically significant, and not merely a consequence of experimental inconsistency. Upon examination of the LSD values calculated for each of the K1 and K8 populations, evidence for real underlying differences was obtained. This suggested that genetic factors were influencing the susceptibility/resistance to beetle herbivory to some degree. For example, in K1, the difference between the mean values for individual 776 (64.48 mm^2 leaf area remaining) and individual 93 (107.87 mm^2 leaf area remaining) was approximately three-fold greater than the associated LSD (13.353). Likewise in K8, the difference between individuals 776 (75.50 mm^2 leaf area remaining) and 132 (107.87 mm^2 leaf area remaining) was approximately three-fold greater than the associated LSD (10.213). These data suggest that, with respect to variation in susceptibility to beetle feeding, either K1 or K8 would be a good choice.

4.5 Chapter summary

- This Chapter describes the establishment of a large segregating mapping population as a basis for genetic mapping of agronomic importance in willow.
- A new population was required, as the existing K3 mapping population did not appear to segregate as required for all traits to be targeted in the study. Furthermore, only 66 progeny were available at this time.
- The progeny of two crosses were chosen initially as candidates to increase the likelihood that segregation for all traits of interest would be achieved. A single population was then chosen later on the basis of early trait assessments.
- The two candidate crosses were selected from seven potentially suitable crosses made as part of the European Willow Breeding Programme. These crosses were deemed likely to segregate for traits of interest and had all produced adequate numbers of seed to produce a sufficiently large mapping population to underpin QTL analysis.
- The pedigrees of all the potential parents of the crosses were checked using AFLP markers. A discrepancy in the pedigrees of three crosses was identified, in turn, ruling them as potential mapping populations
- Two populations derived from crosses between different full-sib progeny of the K477 population (produced in Sweden) were elected as the mapping populations that were established for initial study. This choice was made on the basis of expected trait segregation and seed availability.
- These two full-sib populations, named K1 and K8, were planted in a field trial designed specifically to allow the collection of trait data suitable for use in later QTL analysis.
- Following establishment, field-based trait assessments were performed on both populations for yield-related traits and field-based resistance to *Melampsora* rusts during the first year of growth. Laboratory assays of rust resistance and palatability to the blue willow beetle (*P. vulgatissima*) were also performed.
- For the majority of traits assessed, more even distribution of trait values across the observed ranges was observed in the K8 population compared to K1. Therefore, this full-sib population, comprising 947 progeny, was chosen for more detailed phenotypic study and to underpin later QTL analyses.

Chapter 5. Construction of the K8 linkage map

5.1 INTRODUCTION 154

5.1.1 General strategy 154

5.1.2 Pedigree authentification 155

5.2 MATERIALS AND METHODS 157

5.2.1 Plant material 157

5.2.2 High-throughput DNA extraction from K8 progeny 157

5.2.3 Identification of the correct K8 parents 157

5.2.4 Development of additional SSR markers 158

5.2.5 Testing and screening of SSR primer sets 158

5.2.5.1 Screening microsatellites for detection by autoradiography 159

5.2.5.2 Screening microsatellites for fluorescent detection 159

5.2.6 Development of microsatellite multiplexes 160

5.2.7 Screening microsatellite multiplexes 161

5.2.8 Scoring microsatellite markers 163

5.2.9 Bulked segregant analysis 163

5.2.10 AFLP screening 165

5.2.11 Map construction 165

5.3 RESULTS 166

5.3.1 Identification of the K8 parents 166

5.3.2 Development of additional microsatellite markers 168

5.3.3 Testing for microsatellite polymorphisms in K8 169

5.3.4 Development of microsatellite mutiplexes 169

5.3.5 Microsatellite screening of the K8 population 172

5.3.6 Bulked segregant analysis 174

5.3.7 AFLP markers 177

5.3.8 Construction of the parental maps 178

5.3.9 The K8 consensus map 182

5.3.10 Mapping AFLP markers putatively linked to microsatellite SB226 183

5.3.11 Identification of homeologous K3 and K8 linkage groups 189

5.4 DISCUSSION 194

5.4.1 Identification of the K8 parents 194

5.4.2 Microsatellite markers 195

5.4.2.1 Microsatellite marker development 195

5.4.2.2 Microsatellite screening 196

5.4.3 AFLP markers 197

5.4.4 Bulked segregant analysis 198

5.4.5 The K8 parental maps 199

5.4.6 The K8 consensus map 200

5.4.7 Comparisons with the K3 map 201

5.5 CHAPTER SUMMARY 202

5.1 Introduction

This chapter describes the construction of a genetic linkage map based on the large-scale mapping population K8 (Chapter 4). This map will subsequently be used for QTL analyses aimed at the identification of genomic regions underlying traits of agronomic importance in willow (Chapter 6) that are shown to segregate in this population (Chapter 4).

5.1.1 General strategy

In order to maximise the chances of detecting significant associations between markers and QTL, a large mapping population of 947 progeny was generated in K8 (Chapter 4). Genotyping large numbers of mapping progeny is expensive and labour-intensive and it would be uneconomical, both in terms of time and cost, to embark on a large-scale genotyping programme if the intended objectives may be achieved using a smaller sample size. Consequently, although phenotypic measurements were performed on all 947 progeny (Chapter 4), initial linkage studies were based on genotyping only 480 individuals (plants 1-480). When compared to analogous studies in other outbreeding species, this sub-population still represents a relatively large number of segregating offspring for inclusion in linkage mapping and QTL analyses. For example, a QTL study of bud flush and bud set in poplar was based on two full-sib populations comprising 55 and 346 individuals, respectively (Frewen *et al.*, 2000). If necessary, for high-resolution mapping and more accurate QTL analysis in future, the number of individuals can be extended to include the entire K8 population.

For map construction, emphasis was placed on mapping microsatellite markers as these can be more highly informative in both linkage and QTL analyses in outbreeding populations compared with dominant markers such as AFLPs (Section 3.4.1). Furthermore, if sufficient microsatellite markers were shown to segregate in both the K3 and K8 mapping populations, then alignment and future comparative analysis of corresponding linkage groups in the two maps would be possible. However, given the relatively large number of progeny to be screened, and limitations regarding the time available for map construction within the project, calculation of a map based entirely on microsatellite markers would have been too ambitious an undertaking. The fact that microsatellite markers are conventionally screened in single locus assays and that only

a limited number of markers of this type available at the start of the project would also have hindered construction of a map based purely on microsatellite loci. To address these potential bottlenecks, a large number of additional microsatellite markers were developed and the use of multi-locus microsatellite screening protocols based on fluorescent multiplex systems were examined. However, the latter only became possible with the acquisition of an ABI 3100 Genetic Analyser during the last 10 months of the project whilst initial microsatellite genotyping assays were run using autoradiographic techniques

In addition to microsatellite markers, several AFLPs were mapped. For more informed selection of AFLP primer combinations for mapping, bulked segregant analysis (BSA) was performed prior to AFLP genotyping in order to identify those combinations that would yield markers linked to microsatellite locus SB226. This locus was one of the first microsatellites to be mapped on K8, and was identified in early non-parametric QTL mapping screens as linked to a locus involved in field-based rust resistance (see Section 6.4.3 for details). Saturation of this map region with additional molecular markers should facilitate more detailed analysis of this region in future work. BSA was also performed on DNA pools of rust-resistant and rust-susceptible phenotypes to identify additional markers that segregated with the rust-resistant phenotype.

Map construction was performed according to the double pseudo-testcross strategy as employed earlier for calculation of the K3 map (Section 3.1.2), i.e. separate parental maps were generated prior to construction of a consensus map.

5.1.2 Pedigree authentification

The first marker to be screened against the K8 progeny was a co-dominant ESTP marker (ESTP1; Section 3.2.10). Segregation at this locus revealed a discrepancy between the progeny and parents. The expected male (K477-R11) and female (K477-S1) parents appeared homozygous and heterozygous for this marker, respectively. Therefore, 1:1 Mendelian segregation of homozygotes:heterozygotes would have been expected in the progeny. However, all were homozygous at this locus indicating that K477-R11 could not have been a parent of the mapping population. At a second locus, microsatellite SB54, both parents were heterozygous for this marker while the progeny were all homozygotes. This indicated that neither of the expected parents could have

been used in the original K8 cross. However, no alleles were detected in the progeny that were not detected in the expected parents, suggesting that the true parents may be relatives of those expected. In response, a panel of available microsatellite markers were used to identify the true parents of the mapping population from likely candidates, i.e. the 21 full-sibs of the K477 population from which the expected parents were derived.

5.2 Materials and Methods

5.2.1 Plant material

The K8 mapping population was established and grown as described in Chapter 4. After coppicing at the end of first year of growth, fresh leaf shoots were collected from re-generating stools during the early part of the 2001 growing season. Leaf material was directly placed into 96-well format deep-well collection tubes supplied as part of the DNeasy® 96 Plant DNA Extraction Kit (Qiagen). Care was taken to sample from stool '3' (Figure 4.1) in each plot of six replicates for all individuals. In cases where this stool had not established, stool '4' was used as an alternative. Leaf material was stored at –20°C for a maximum of 24 hours prior to DNA extraction.

5.2.2 High-throughput DNA extraction from K8 progeny

Immediately prior to DNA extraction, 96-well format collection tubes (Qiagen) containing leaf samples were frozen in liquid nitrogen and a MM300 Mixer Mill (Retsch) was used to grind the tissue into a fine powder. DNA extractions were then performed using the DNeasy® 96 Plant DNA Extraction Kit (Qiagen) according to the Manufacturers' instructions. DNA quality and yield were assessed by agarose gel electrophoresis (Section 2.4.1 and 2.4.2). DNA extracts were diluted to a standard concentration of 10 ng/µl, with SDW, and aliquots were transferred to 96-well U-bottom plates (Matrix Technologies) for use in marker screening. Where plants had failed to establish in the field and, therefore, could not be sampled, collection tubes were left empty. In this way, the order of the samples was maintained and the blanks could be used as a negative control for DNA extractions.

5.2.3 Identification of the correct K8 parents

Genomic DNA was extracted from fresh leaf material collected from the 21 inviduals of the K477 population using the Nucleon™ Phytopure™ Genomic DNA Extraction Kit as described in Section 2.4.1. Microsatellite primer sets SB24, SB38, SB54, SB85, SB93, SB194, SB226, SB276, SB293, SB354, SB525 and SB565 were each used to screen the K477 population and 12 randomly selected K8 progeny, using the

radioisotope-based protocol as described in Section 2.5.1.4.1. Allele segregation at each locus was examined to allow systematic exclusion of K477 individuals that could not be parents of the mapping population.

5.2.4 Development of additional SSR markers

For development of additional microsatellite markers, more microsatellite library inserts (Section 3.2.5) were characterised. Plasmid DNA was isolated as described in Section 2.5.1.2 and sequenced in a single direction using the M13 Universal forward primer according to Section 2.4.4. Vector and adapter sequence that flanked the willow DNA insert (see Edwards *et al.*, 1996) was manually identified and removed. Edited sequences were then screened for duplications against previously generated sequence data using the Sequencher™ software package (Section 2.4.5).

Primer sets were designed to microsatellite flanking regions (Section 2.5.1.3) if: 1) sufficient sequence length was available between the vector/adapter and the microsatellite motif; 2) successful primers had not already been designed to a homologous library insert; 3) the nucleotide composition of sequence flanking the microsatellite was suitable for primer design. Primers were synthesised and desalted by Sigma Genosys Ltd.. For several duplicates to which primer sets had previously been designed but had not generated a PCR product, alternative primers were selected. All primer sets were designed for use under the same PCR conditions in an attempt to aid subsequent development of PCR multiplexes.

5.2.5 Testing and screening of SSR primer sets

Potentially useful microsatellite primer sets previously developed for mapping in K3 (Section 3.3.2) and those designed as part of the work described in this chapter were tested for their ability to detect polymorphic PCR products. Genomic DNA of the parents and grandparents of the K8 population was used with the experimental conditions described in Section 2.5.1.4.1. Those that successfully amplified easy-to-score, polymorphic products between the parents were then used to screen genomic DNA from the progeny of the K8 mapping population, using one of two visualisation methods: 1) microsatellite PCR products were amplified using radio-labelled PCR primers and visualised by autoradiography (Section 2.5.1.4.1); or 2) PCR products

were generated using fluorescently-labelled primers for detection on an ABI 3100 Genetic Analyser (section 5.2.5.2). DNA extraction control samples (Section 5.2.2) were also used as a negative control in PCR amplifications.

5.2.5.1 Screening microsatellites for detection by autoradiography

PCR amplification of microsatellite alleles was performed as described in section 2.5.1.4.1 and resulting products were detected using denaturing polyacrylamide gel electrophoresis and autoradiography (Section 2.5.1.4.1). Multiple loadings on acrylamide gels were performed where applicable, i.e. for loci where allele sizes were sufficiently similar, the resolution of multiple 'rows' of alleles on a single gel was possible (see Figure 5.4 for an example).

5.2.5.2 Screening microsatellites for fluorescent detection

Microsatellite loci were amplified in PCR reactions containing a newly synthesised fluorophore-labelled forward primer and an unlabelled reverse primer. Each forward primer was end-labelled with one of four commercially available fluorescent labels: 6-FAM™ (blue), VIC™ (green), PET™ (red) or NED™ (yellow). All fluorescently-labelled primers were synthesised by Applied Biosystems.

To determine microsatellite allele sizes, each locus was amplified separately. PCR reaction components and thermocycling conditions were identical to those described in section 2.5.1.4.1, except that fluorescently-labelled forward primer was substituted for γ-^{33}P-labelled forward primer. Also, the increased sensitivity of the fluorescent detection system allowed for a reduction of the number of PCR cycles to 30. Following amplification, PCR products were resolved on an ABI 3100 Genetic Analyser (Section 5.2.7) and genotypes were determined using GeneMapper® software (Applied Biosystems).

5.2.6 Development of microsatellite multiplexes

To increase the efficiency of microsatellite genotyping of the K8 population, with regard to both time and expense, microsatellite multiplex systems were developed. Two complementary strategies were employed: 1) multiple microsatellite primer sets were included in a single PCR reaction to allow simultaneous amplification of several distinct loci; 2) amplification products from PCR multiplexes were pooled prior to genotype detection.

Initial selection of loci to be included in a given multiplex was made on the basis of relative size differences between alleles detected during primer testing (Section 5.2.5). If two microsatellite primer sets produced well-defined alleles of sufficiently different sizes, primers for both loci were labelled with the same fluorophore. Where different loci generated overlapping alleles, primers were labelled with different fluorophores. In such cases, primer labels were selected on the basis of their different emission spectra to prevent scoring problems associated with matrix 'pull up' (Applied Biosystems) when signals are strong.

PCR multiplexes were then tested using the conditions described in Section 5.2.7, with the exception that multiple primer sets were included in a single reaction. The relative concentrations of each primer set was based on the strength of signal initially observed visually after autoradiography, i.e. microsatellite primer sets that generated products with intense signals during initial tests were included in multiplex PCR at lower concentrations relative to those primer sets that generated weak signals.

To determine whether all expected alleles could be reliably detected, suitable microsatellite PCR multiplexes were tested against a panel of eight randomly selected K8 individuals. In cases where all expected alleles were observed, but the signal strength associated with one or more included loci was weak, a second round of optimisation was performed in which primer ratios were adjusted further to obtain a more uniform signal strength. Where loci completely failed to generate a product in a given PCR multiplex, the locus was removed from the multiplex and tested in combination with other suitable loci. When successful combinations were identified, resulting allele sizes were recorded and the possibility of pooling products from different PCR multiplexes for simultaneous electrophoretic resolution was examined.

5.2.7 Screening microsatellite multiplexes

Once definitive PCR and/or loading multiplexes were identified (Table 5.1), genotyping of K8 progeny was performed. PCR reactions were set up as described in Table 5.1 and thermocycling performed as described in Section 5.2.5.2. For loading multiplexes, PCR products were mixed in equal volumes and 0.5 µl of the resulting mixture used in a 10 µl loading cocktail containing Hi-DI ™ Formamide (Applied Biosystems) and Genescan LIZ™-500 size standard (Applied Biosystems) according to the Manufacturer's instructions.

Table 5.1. Microsatellite multiplexes used for K8 progeny screening. All figures represent µl volumes used in PCR reactions. Colours refer to fluorescent labels.

Loading Multiplex	PCR multiplex	Locus	F&R primers (100 ng/µl)	DNA (2.5 ng/ul)	10x PCR Buffer (Invitrogen)	$MgCl_2$ 50 mM	dNTP (5 mM)	SDW	*Taq* polymerase 5U / µl (Invitrogen)	Final volume
1	1	SB565	0.2	2	1.25	0.375	0.5	6.175	0.1	12.5
		SB880	0.2							
		SB896	0.15							
		SB784	0.25							
		SB869	0.25							
2	2	SB288	0.2	2	1.25	0.375	0.5	6.075	0.1	12.5
		SB306	0.2							
		SB955	0.2							
		SB194	0.3							
		SB293	0.2							
A	A1	SB111	0.12	1	1.25	0.375	0.5	7.755	0.1	12.5
		SB196	0.22							
		SB913	0.2							
		SB945	0.22							
	A2	SB55	0.2	1	1.25	0.375	0.5		0.1	12.5
		SB80	0.2							
		SB405	0.15							
C	C1	SB51	0.2	1	1.25	0.375	0.5		0.1	12.5
		SB88	0.2							
		SB918	0.25							
		SB276	0.25							

Table 5.1 continued.

Loading Multiplex	PCR multiplex	Locus	F&R primers (100 ng/µl)	DNA (2.5 ng/ul)	10x PCR Buffer	$MgCl_2$ 50 mM	dNTP (5 mM)	SDW	*Taq* polymerase 5U / µl	Final volume
	C2	SB201	0.36	1	1.25	0.375	0.5		0.1	12.5
		SB226	0.21							
		SB514b	0.17							
D	D1	SB274	0.26	1	1.25	0.375	0.5		0.1	12.5
		SB24	0.27							
		SB522b	0.2							
		SB430	0.2							
	D2	SB488	0.2	1	1.25	0.375	0.5		0.1	12.5
		SB126	0.25							
		SB93	0.26							
		SB354	0.17							
E	E1	SB1035	0.18	1	1.25	0.375	0.5	7.815	0.1	12.5
		SB1060	0.23							
		SB1084	0.1							
		SB1045	0.22							
	E2	SB1092	0.2	1	1.25	0.375	0.5	7.295	0.1	12.5
		SB1048	0.24							
		SB1308	0.18							
		SB1317	0.15							
		SB988	0.22							
	E3	SB1332	0.25	1	1.25	0.375	0.5	7.795	0.1	12.5
		SB1254	0.18							
		SB1147	0.15							
		SB1249	0.16							
	E4	SB1229	0.15	1	1.25	0.375	0.5	8.315	0.1	12.5
		SB1094	0.22							
		SB1148	0.11							
F	F1	SB914	0.18	0.8	1	0.3	0.4	5.92	0.08	10
		SB504	0.35							
		SB355	0.22							
	F2	SB984	0.2	0.8	1	0.3	0.4	5.62	0.08	10
		SB596	0.3							
		SB331	0.4							
	F3	SB1185	0.2	0.8	1	0.3	0.4	6.62	0.08	10
		SB1318	0.2							
	F4	SB1075	0.24	0.8	1	0.3	0.4	6.7	0.08	10
		SB1196	0.12							

5.2.8 Scoring microsatellite markers

For radioisotope methods, autoradiographs were scored independently by two persons and genotype segregation codes were assigned for each locus in accordance with JoinMap® software requirements (Table 2.1). Electropherograms resulting from fluorescent detection methods were analysed and JoinMap genotype codes assigned using GeneMapper software (Applied Biosystems).

5.2.9 Bulked segregant analysis

Bulked segregant analysis was primarily performed to identify AFLP primer combinations that yielded markers linked to microsatellite locus SB226, which showed linkage to rust resistance in QTL analysis (Section 6.3.4.3). To achieve this, six pools of genomic DNA were generated based on individuals of the K8 population as shown in Table 5.2. In addition, a further two pools were constructed based on ten individuals each with either a field-based rust resistant phenotype or a rust-susceptible phenotype, respectively. Individuals were selected on the basis of the raw data collected in field-based rust assessments performed in 2000 and 2001 (Chapter 4).

Table 5.2. K8 individuals included in DNA pools used for bulked segregant analysis aimed at identifying AFLP markers linked to microsatellite locus SB226.

Bulked Sample	SB226 genotype	K8 progeny included	Number of progeny
1	aa	2, 4, 5, 14, 20, 23, 29, 31, 36, 40	10
2	bb	6, 8, 10, 11, 17, 21, 30, 33, 34, 35	10
3	aa	85, 90, 93, 97, 102, 106, 107, 111, 119, 131	10
4	bb	81, 89, 92, 94, 96, 100, 109, 114, 121, 124	10
5	aa	55, 64, 65, 71, 72, 73, 78	7
6	bb	142, 147, 149, 151, 152, 154, 155	7
Bulked sample	**Phenotype**	**K8 progeny included**	**Number of progeny**
7	Rust resistant	12, 15, 26, 32, 57, 75, 98, 115, 133, 142	10
8	Rust susceptible	77, 78, 87, 106, 117, 120, 123, 131, 164, 230	10

AFLP analysis of bulked DNA samples was essentially performed according to AFLP Method 2 (Section 2.5.2.2). In brief, digestions and ligations, based on *Eco*RI/*Mse*I primer/adapter combinations, were first performed on each constituent DNA sample

separately. Samples were diluted ten-fold in $TE_{0.1}$ and equal amounts taken and pooled according to Table 5.2. To optimise the amount of pooled DNA that should be used in subsequent amplification steps, pre-amplifications were performed on bulked samples 7 and 8 using three different concentrations of template DNA in the PCR reaction (5 µl, 10 µl and 14.9 µl). Following dilution of pre-amplification products, selective amplifications based on primer combinations EACT/MAAG and EAGA/MAAG were performed and products were subsequently resolved by denaturing polyacrylamide gel electrophoresis. Corresponding AFLP products from both K8 parents were also run on the gel to determine whether all loci were detected. Examination of the resulting autoradiographs suggested that there was no discernable difference in bulked profiles when different amounts of DNA template were used and all loci present in the parental profiles were detected. As a result, 10 µl of digested/ligated DNA was routinely used for the pre-amplification stage of subsequent BSA.

All DNA pools were screened using 56 AFLP primer combinations that were available in the laboratory, according to the optimised protocol described above. In addition, K8 parental DNA samples were included in analyses to facilitate identification of particularly informative primer combinations that were detecting large numbers of parental markers. Details of the primer combinations screened are shown in Table 5.3.

Table 5.3. AFLP primer combinations used to screen DNA pools in bulked segregant analysis where '✓' =tested; 'n/t' = not tested.

		*Eco*RI+							
		AGA	AGC	ACT	ACA	ATA	AAC	CAA	CAG
*Mse*I+	AAG	✓	✓	✓	✓	✓	✓	✓	✓
	AAT	✓	✓	✓	✓	✓	✓	✓	✓
	ACC	✓	✓	✓	✓	✓	✓	✓	✓
	ACA	✓	✓	✓	✓	✓	✓	✓	✓
	CCG	✓	✓	n/t	n/t	✓	✓	✓	✓
	CTA	✓	✓	n/t	n/t	✓	✓	✓	✓
	CTG	✓	✓	n/t	n/t	✓	✓	✓	✓
	CTT	✓	✓	n/t	n/t	✓	✓	✓	✓

Following amplification, products were resolved by denaturing polyacrylamide gel electrophoresis with four primer combinations loaded on to each gel. Autoradiographs

were examined for presence of loci that segregated between the respective pools. Markers that were present in bulks 1, 3 and 5, and absent in bulks 2, 4 and 6, or *vice versa*, were labelled as putatively linked to microsatellite locus SB226. Markers present in bulk 7, but not bulk 8, or *vice versa*, were labelled as potentially linked to the field-based rust resistant/susceptible phenotype. Primer combinations that detected such markers in both types of bulk were then preferentially used to screen the K8 progeny for detection of polymorphisms that could be mapped.

5.2.10 AFLP screening

AFLP markers were generated according to AFLP Method 2 (Section 2.5.2.2). PCR products were resolved on denaturing polyacrylamide gels as described in Section 2.5.3, with 80 samples loaded per gel. Following autoradiography, segregating alleles were scored and assigned JoinMap segregation codes according to Table 2.1.

5.2.11 Map construction

Linkage analysis was based on 471 (out of the first 480) individuals of the K8 population. Of the nine that were not included, eight individuals had not survived establishment and one individual (318) was shown not to be a member of the full-sib family.

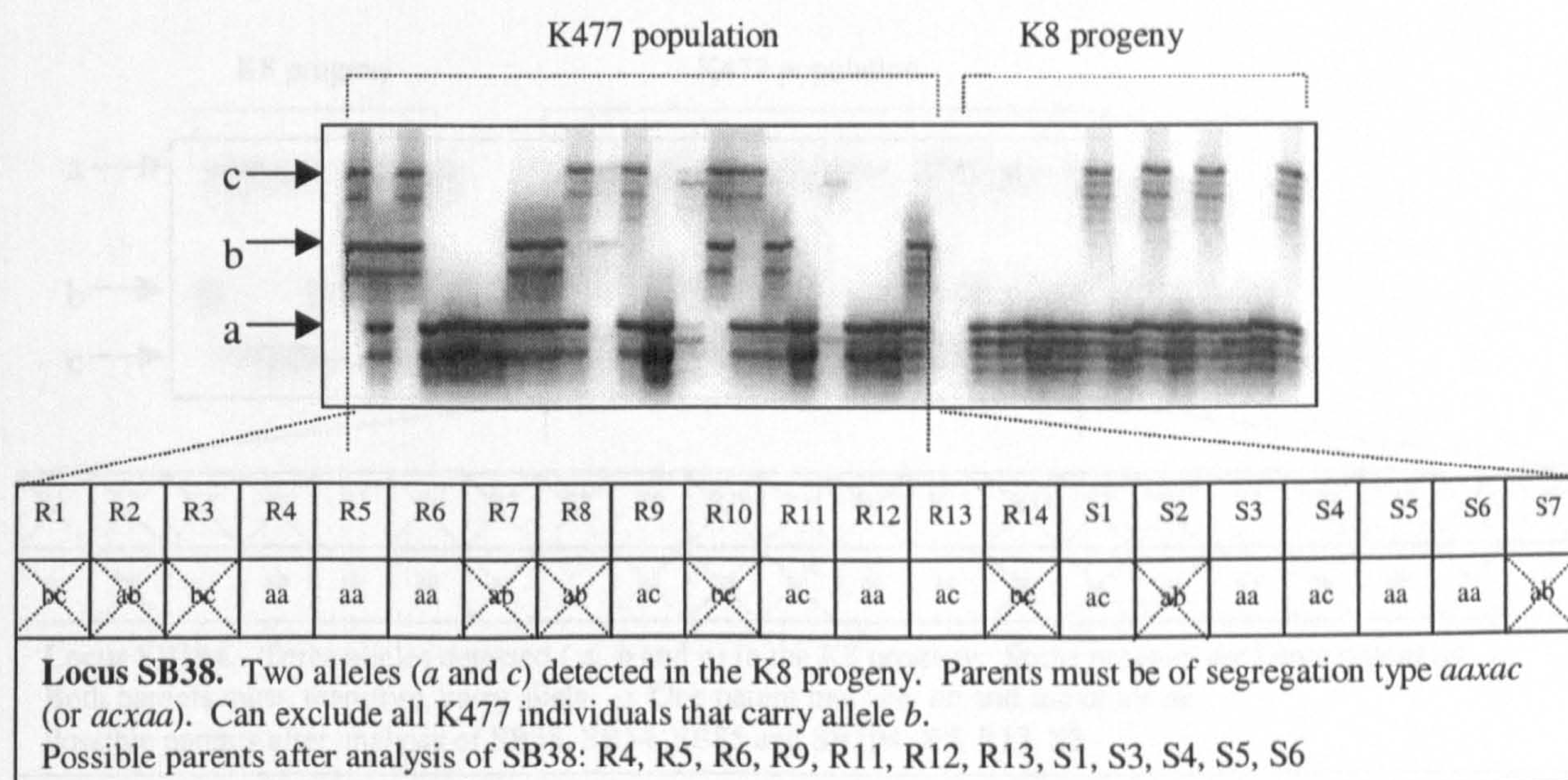
JoinMap version 3.0 software (Van Ooijen and Voorrips, 2001) was used for all aspects of linkage analysis. The software was used to first test markers for segregation distortion using a chi-square test. AFLP markers with highly significant levels of distortion ($P < 0.005$) were excluded from further linkage analysis. Parental and integrated linkage maps were constructed according to the strategy described in section 3.1.2. The Kosambi mapping function was used with the following JoinMap parameter settings unless stated otherwise: Rec = 0.45, LOD = 1.0, Jump = 5. Resulting linkage maps were drawn using MapChart 2.1 software (Voorrips, 2002). Where possible, robustness of the consensus linkage map was tested by comparison of marker orders with separate parental maps.

5.3 Results

5.3.1 Identification of the K8 parents

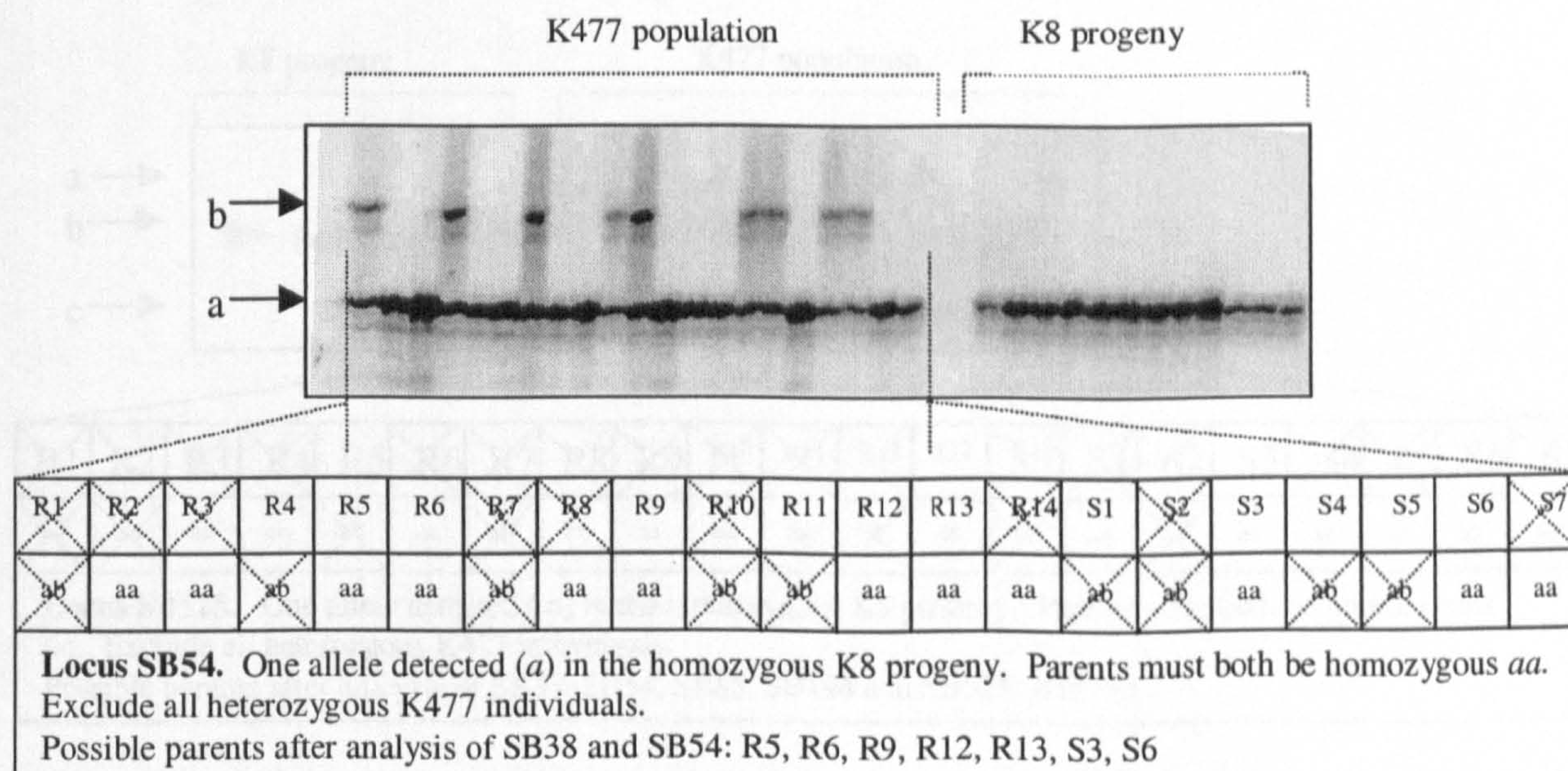
All microsatellites screened against the K477 and K8 populations gave profiles that could be clearly scored. After examination of segregation at five loci, only two K477 individuals, K477-S3 and K477-R13 (♀ and ♂, respectively), remained as possible parents of the K8 mapping population (Figure 5.1).

Figure 5.1. Microsatellite-based exclusion experiment to identify possible parents of the K8 mapping population. Analysis of five loci showed K477-S3 and K477-R13 to be the only possible parents of the mapping population.



R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	S1	S2	S3	S4	S5	S6	S7
bc	ab	bc	aa	aa	aa	ab	ab	ac	bc	ac	aa	ac	bc	ac	ab	aa	ac	aa	aa	ab

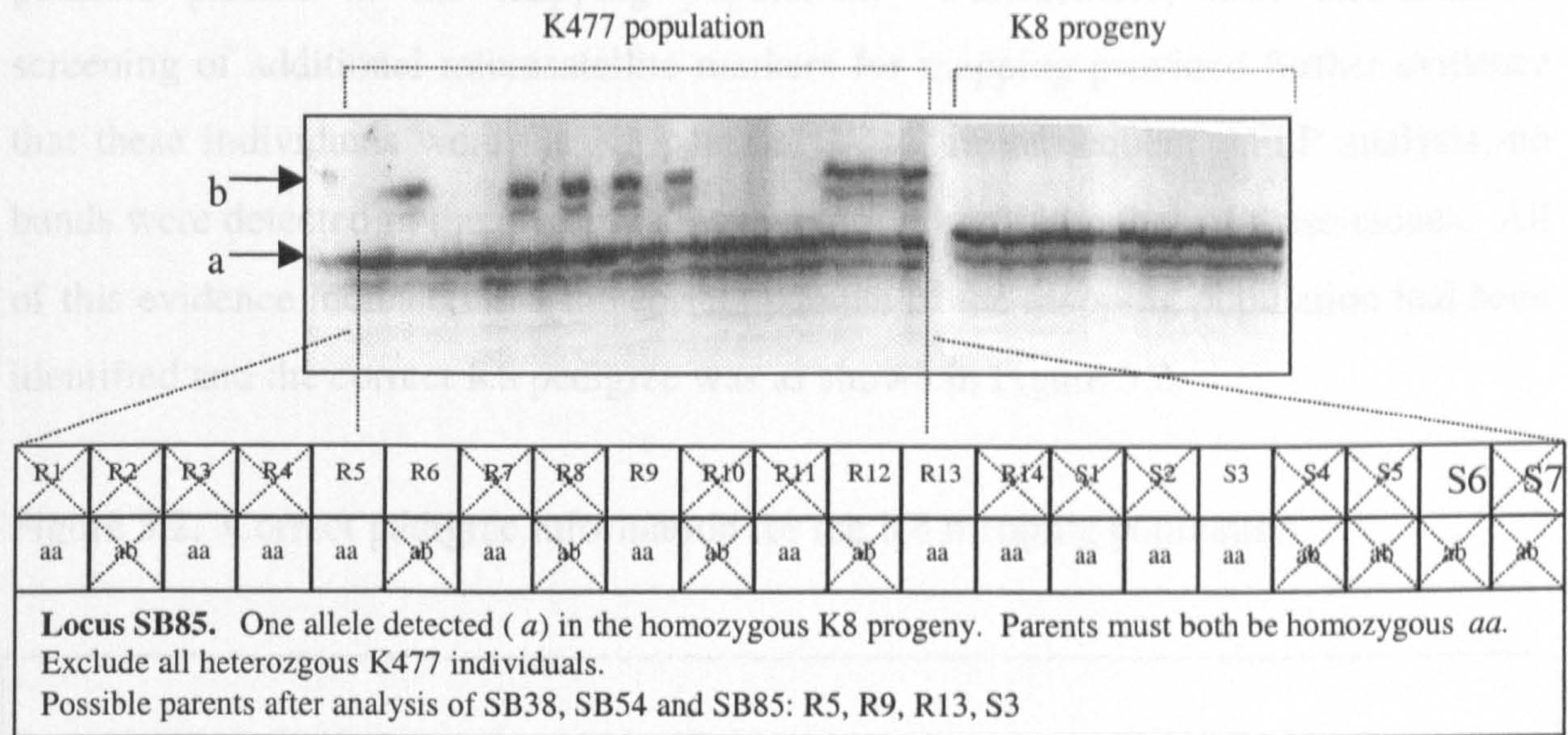
Locus SB38. Two alleles (*a* and *c*) detected in the K8 progeny. Parents must be of segregation type *aaxac* (or *acxaa*). Can exclude all K477 individuals that carry allele *b*.
Possible parents after analysis of SB38: R4, R5, R6, R9, R11, R12, R13, S1, S3, S4, S5, S6



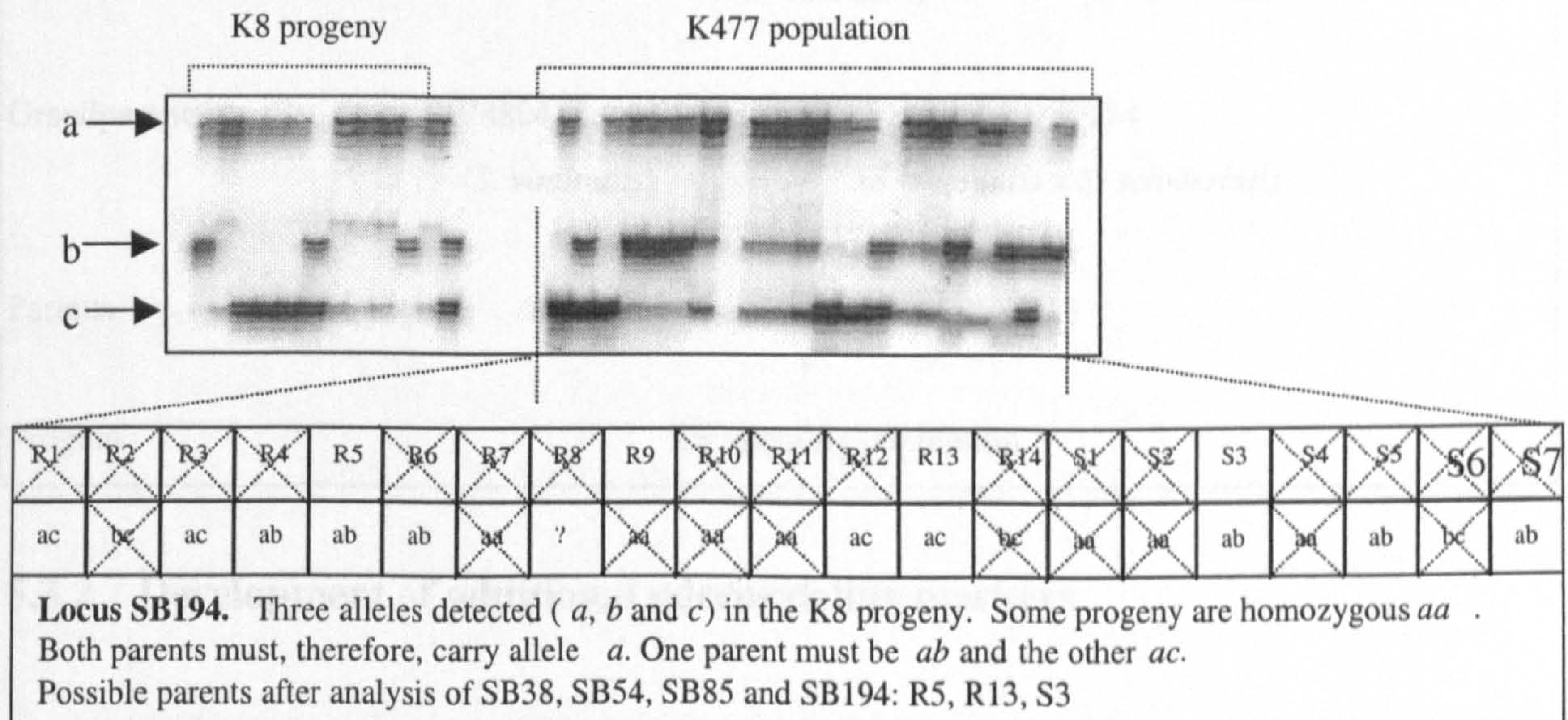
R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	S1	S2	S3	S4	S5	S6	S7
ab	aa	aa	ab	aa	aa	ab	aa	aa	ab	ab	aa	aa	aa	ab	ab	aa	ab	ab	aa	aa

Locus SB54. One allele detected (*a*) in the homozygous K8 progeny. Parents must both be homozygous *aa*. Exclude all heterozygous K477 individuals.
Possible parents after analysis of SB38 and SB54: R5, R6, R9, R12, R13, S3, S6

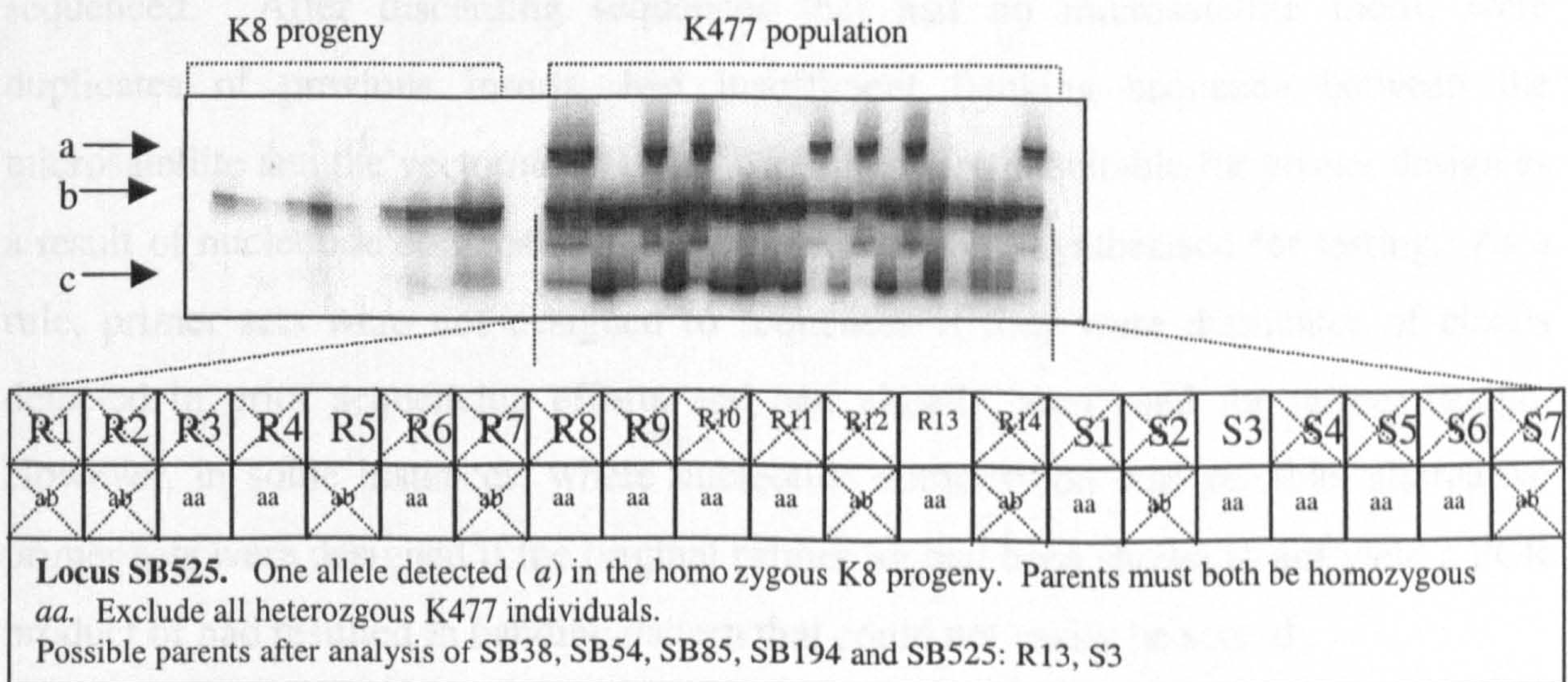
Figure 5.1 continued.



Locus SB85. One allele detected (*a*) in the homozygous K8 progeny. Parents must both be homozygous *aa*. Exclude all heterozgous K477 individuals.
Possible parents after analysis of SB38, SB54 and SB85: R5, R9, R13, S3



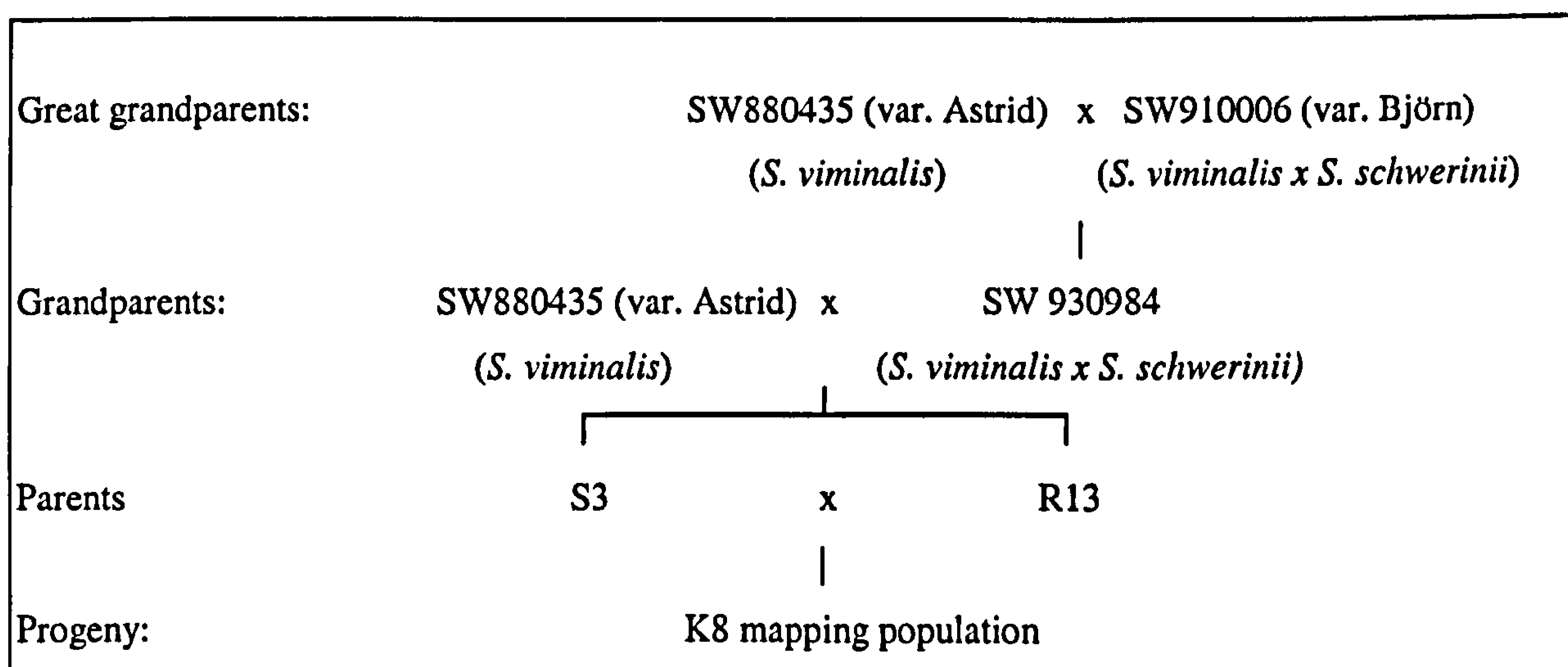
Locus SB194. Three alleles detected (*a*, *b* and *c*) in the K8 progeny. Some progeny are homozygous *aa* .
Both parents must, therefore, carry allele *a*. One parent must be *ab* and the other *ac*.
Possible parents after analysis of SB38, SB54, SB85 and SB194: R5, R13, S3



Locus SB525. One allele detected (*a*) in the homo zygous K8 progeny. Parents must both be homozygous *aa*. Exclude all heterozgous K477 individuals.
Possible parents after analysis of SB38, SB54, SB85, SB194 and SB525: R13, S3

Segregation at the other microsatellite loci tested were in agreement with the results presented in Figure 5.1, in that K477-S3 and K477-R13 were never excluded as the possible parents of the mapping population. Furthermore, later microsatellite screening of additional microsatellite markers for mapping provided further evidence that these individuals were the K8 parents. Also, in subsequent AFLP analysis, no bands were detected in the progeny that were not present in either of these clones. All of this evidence indicated that the correct parents of the mapping population had been identified and the correct K8 pedigree was as shown in Figure 5.2.

Figure 5.2. Correct pedigree information for the K8 mapping population.



5.3.2 Development of additional microsatellite markers

Willow DNA inserts from a further 398 microsatellite library inserts were successfully sequenced. After discarding sequences that had no microsatellite motif, were duplicates of previous inserts, had insufficient flanking sequence between the microsatellite and the vector/adapter, or were otherwise unsuitable for primer design as a result of nucleotide composition, 183 primer sets were synthesised for testing. As a rule, primer sets were not designed to sequences if they were duplicates of clones detected in prior sequencing efforts and had already been used for primer design. However, in some instances, where nucleotide composition was suitable, alternative primer sets were designed if the original primer set had been shown to not yield a PCR product or had resulted in banding pattern that could not easily be scored.

5.3.3 Testing for microsatellite polymorphisms in K8

In total, 349 microsatellite primer sets developed either during the EU-AIR project, during the K3 mapping phase of this project or in the latest marker development effort, were tested for their ability to detect polymorphism between the parents and grandparents of the K8 mapping population. Single locus, easy-to-score, polymorphic PCR products were observed for 71 (~20%) of the loci when amplified under standard PCR testing conditions (Section 2.5.1.4.1.).

Of the 56 microsatellite loci included in linkage analysis of the K3 population (Chapter 3), 52 were tested for informativeness in K8. Segregating alleles that could be reliably scored were observed with 29 microsatellite primer sets. Twelve loci were monomorphic and 11 primer sets generated complex banding patterns that could not be scored.

Primer re-design to duplicated inserts for which original primer sets did not yield scorable products was successful in some instances. For example, locus SB1366 was homologous to the duplicate loci SB886 and SB890. While primers designed to the latter two inserts failed to generate scorable products when tested, alternative primers designed to SB1366 were successful in yielding a banding pattern that was easily interpreted. However, this was not the case for all duplicated loci, and in several cases primer re-design proved unsuccessful, e.g. three different primer sets generated for the homologous inserts SB321, SB606 and SB1300 did not yield scorable products.

5.3.4 Development of microsatellite multiplexes

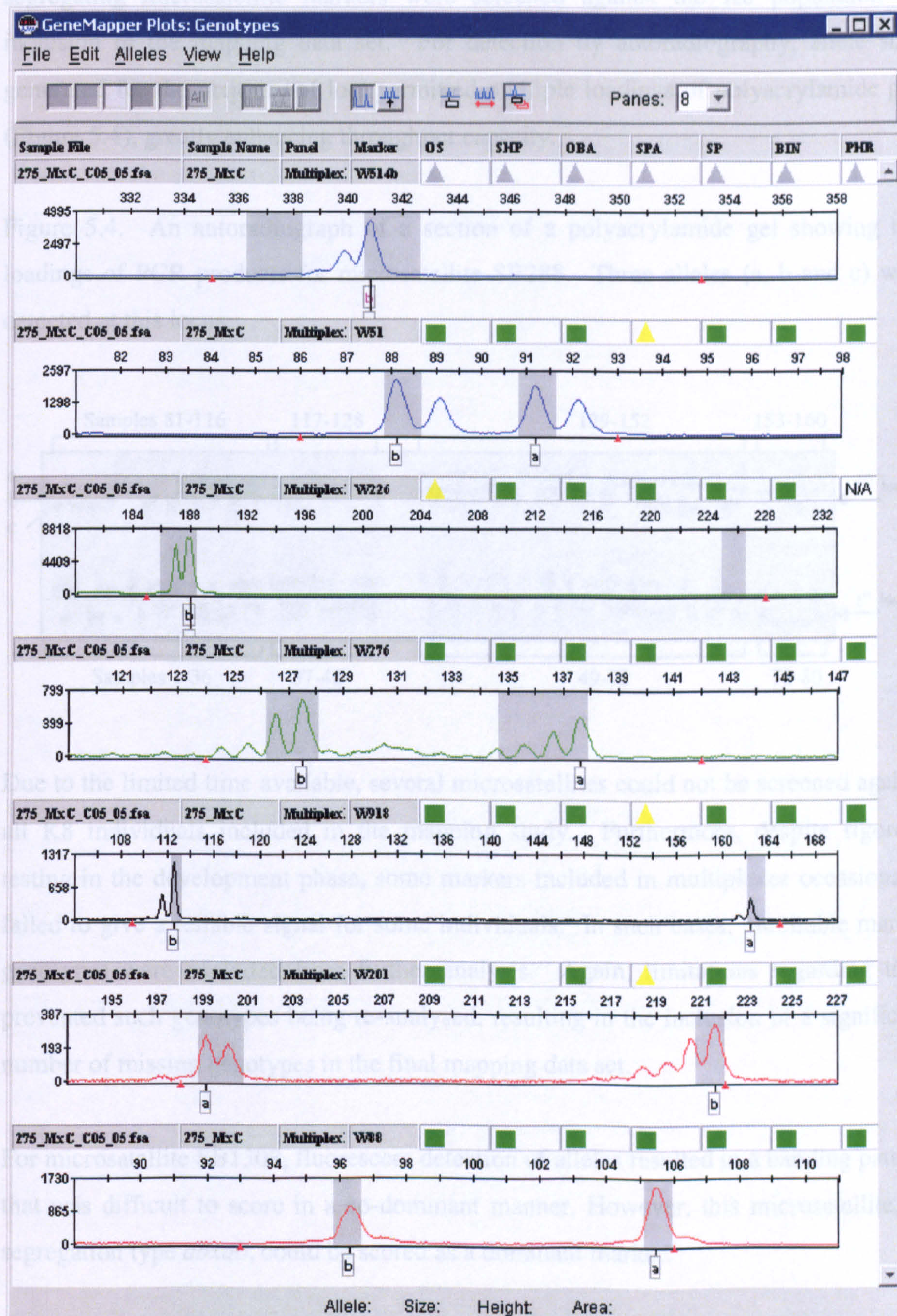
In some cases, microsatellite PCR multiplex design proved straightforward with originally selected locus combinations working well without the need for further optimisation, e.g. multiplexes 1 and 2 (Table 5.1). However, for the remaining multiplexes, development was more difficult and a significant amount of time was invested in identification of successful locus combinations. This was deemed worthwhile due to the increased throughput that could potentially be achieved. Several microsatellites, including SB233, SB268, SB420 and SB587 failed to generate scorable products in any of the multiplex combinations tested. Efforts to incorporate

such loci in alternative multiplexes were aborted if the markers could be more efficiently mapped by more conventional, single-locus, assays.

The number of microsatellite loci successfully amplified in a single multiplex PCR reaction ranged from two to five. By pooling PCR products from several PCR multiplexes prior to loading on the ABI 3100 Genetic Analyser, it was possible to interrogate from five to sixteen loci in a single capillary injection.

An example of data obtained from successful microsatellite multiplexing strategy is shown in Figure 5.3. The figure comprises a screenshot taken from GeneMapper software, illustrating the genotype scores obtained for K8 individual 275 at microsatellite loci SB51, SB201, SB226, SB276, SB514b and SB918 (Multiplex C: Table 5.1).

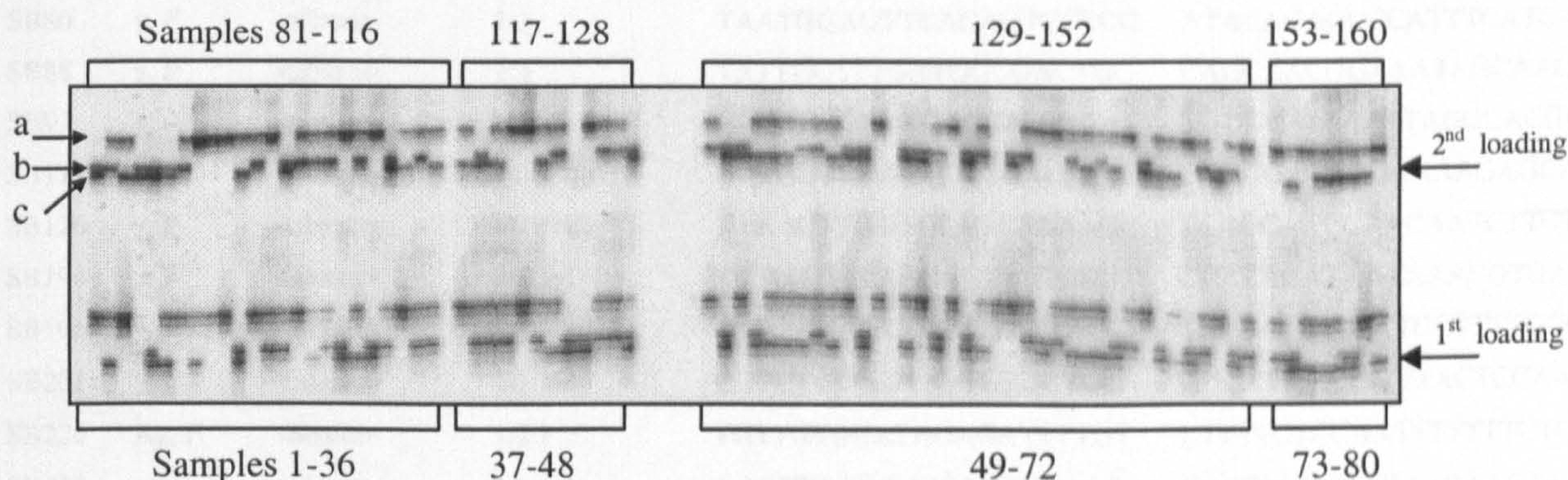
Figure 5.3. An example of successful genotype detection using a PCR- and loading-based multiplexing strategy.



5.3.5 Microsatellite screening of the K8 population

Using both radioisotope and fluorescence-based detection methods, a total of 71 segregating microsatellite markers were screened against the K8 population for inclusion in the mapping data set. For detection by autoradiography, allele sizes generated for the majority of loci permitted multiple loadings of polyacrylamide gels (Figure 5.4), greatly enhancing throughput capacity.

Figure 5.4. An autoradiograph of a section of a polyacrylamide gel showing two loadings of PCR products for microsatellite SB288. Three alleles (a, b and c) were detected at this locus.



Due to the limited time available, several microsatellites could not be screened against all K8 individuals included in the mapping study. Furthermore, despite rigorous testing in the development phase, some markers included in multiplexes occasionally failed to give a reliable signal for some individuals. In such cases, unreliable marker genotypes were excluded from further analysis. Again, limitations regarding time prevented such genotypes being re-analysed, resulting in the inclusion of a significant number of missing genotypes in the final mapping data set.

For microsatellite SB1308, fluorescent detection of alleles resulted in a banding pattern that was difficult to score in a co-dominant manner. However, this microsatellite, of segregation type *abxab*, could be scored as a dominant marker.

Early microsatellite screening identified K8 individual 318 as possessing several alleles that were not present in the K8 parents. Comparisons of these alleles against marker data generated in previous studies (S.J. Hanley; J.H.A. Barker; unpublished

results) suggested that DNA sample for this individual had originated from the biomass variety Jorr, also included in the field trial as a reference clone (Section 4.1.3). All genotype data generated from this DNA sample was, therefore, discarded.

Details of microsatellite loci included in K8 linkage analysis are shown in Table 5.4.

Table 5.4. Details of microsatellite loci included in the K8 mapping data set.

Locus	Detection method	Segregation type (♀ x ♂)	Expected segregation ratio	Forward primer (5'→3')	Reverse primer (5'→3')
SB24	F, γ	<abxac>	1:1:1:1	ACTTCAATCTCTCTGTATTCT	CTATTTATGGGTTGGTCGATC
SB38	Ag	<aaxab>	1:1	CCACTTGAGGAGTGTAAGGAT	CTTAAATGTAAAACTGAATCT
SB51	γ, F	<abxaa>	1:1	GGTAATAGCTTGAGCCTTCAT	AACTTTGCTCTGGTCCTCTTT
SB55	γ, F	<abxab>	1:2:1	TAGGGAATCCTAGTTCTTGCA	CTTCAAAGACCCTAACTTGAA
SB80	γ, F	<abxaa>	1:1	TAATGGAGTTCACAGTCCTCC	ATACAGAGCCCATTTCATCAC
SB88	γ, F	<abxaa>	1:1	TATTGCTTTGATGGCGACTGC	CAGCAACGGAAATAGCAACAG
SB93	γ, F	<aaxab>	1:1	GACGCACATACACCATTACAC	TGTTAGAAAATTAGGCACGGA
SB111	γ, F	<abxab>	1:2:1	GATAGTCAAGGTTGGAGATGG	GGTGGAGAAGAAAAGAGCAGA
SB126	γ, F	<abxac>	1:1:1:1	TAAACTTGTGTTAGTGAAAGC	TCAGCAACCAACAATCTTCTC
SB194	γ, F	<abxac>	1:1:1:1	TGTGAGATAAGATTTGTCGGT	CCATAAATAAAAAACGTGAAC
SB196	γ, F	<aaxab>	1:1	CTGTTTCCTGCCACTATTACC	TATAATCTGTCTCCTTTTGGC
SB201	γ, F	<aaxab>	1:1	CCTCTTTTTCTATTGTGGTCT	GGCATGTATTTTTACTCCAAC
SB226	Ag, F	<abxab>	1:2:1	TGTTGTGCATAGAGATTTTGT	CTTTTCTCCAATTTTTTTCTG
SB233	γ	<abxac>	1:1:1:1	AAATTACCGTCCAACTAAAGA	CATTAGCCATGAACAAGTAAA
SB268	γ	<abxaa>	1:1	CTAGTTCTGGTGGGGAAGATG	TTCTAAACCATCATTTGGGTG
SB274	γ, F	<abxac>	1:1:1:1	CCGACCCTCTTTTCTTCTATC	AGCACAGATTGAATGATTGAG
SB276	γ, F	<abxaa>	1:1	AATAACCACTCCATCTTTCAC	TATTAGTGTTTGCTTTGGAGC
SB288	γ, F	<abxac>	1:1:1:1	AGGCTTCACTGTCTCCTCTAG	TCATCACAGCATCTTATCAGG
SB293	γ, F	<abxab>	1:2:1	TGATTGGGCTAAAGATGAAGC	AACTCAGCAACCACCAGAAAC
SB306	F	<abxab>	1:2:1	GATAAGCATCTGTCGATAGGC	GGTTATCTTCATTCATGTATC
SB331	γ	<aaxab>	1:1	TGTTTGTTTGGGAGTATGTTG	TGAATCCTATCATTTGGCAAC
SB337	γ	<aaxab>	1:1	TACTCTGTCTTTTGCATTTCC	ATACCCATTAGAGAAGGTTTC
SB354	γ, F	<abxaa>	1:1	TGTCTTTGTGTAACGACTCTG	ATCCATATAGCTGATCCCAAC
SB355	γ, F	<abxab>	1:2:1	TTCATTTTCATCTCGAAGCAC	ATTTAGTGCGGATCTTTCAAC
SB405	γ, F	<abxaa>	1:1	TGACAACGCCATTCACATGAC	TTCCTTTCTTCATCAGTAGCC
SB420	γ	<abxab>	1:2:1	TGATGAGGTAAAAAGATGGAG	TAAGGAACACATAATTGAGGG
SB430	γ, F	<abxaa>	1:1	CACCCTCATAACAAAAATGGC	CAAATCAGAAAAGAAGTTAAC
SB488	γ, F	<abxab>	1:2:1	AGCGCAAAGAGAATCGACAAC	TAAGTGTTCCAAGAAGTTGCC
SB504	γ, F	<aaxab>	1:1	TGTTGATTAAGATGTGACTGG	ATCATGCCACTAAACACAACC
SB514b	γ, F	<abxab>	1:2:1	CTTCATCCTGACCAAAGTTAC	AAGGGAGATTCATTACCAGAG
SB522b	γ, F	<aaxab>	1:1	CTGGAGTGAGAATGGTGTATC	GTGGACTAACGGATCTATCAC
SB532	γ	<abxab>	1:2:1	TTGCCTTTCTTGTGGTCAAAC	ATCTTCGCTCCTTACTTTCTG
SB565	γ, F	<abxab>	1:2:1	GAAAATATAATGCCCAGGAAG	ACAGAACACAGCGACATGAAC
SB587	γ	<abxac>	1:1:1:1	CAATTCATCCATTAGCACCTG	AAGAAACACGGGACAAACTAC
SB784	γ, F	<aaxab>	1:1	GCACAGATAAAAAATTGGTTG	ATATGACTAGGAGGATGTGTT
SB869	γ, F	<aaxab>	1:1	AATTCTAAAGCCACTTGTGAG	AGATTTTTCGTGGTGTGTGTG
SB880	γ, F	<abxab>	1:2:1	AAAACACCAGAGAACTGCTAC	TACAACTTCATCTTCCTCTCC

Table 5.4 continued.

Locus	Detection method	Segregation type (♀ x ♂)	Expected segregation ratio	Forward primer (5'→3')	Reverse primer (5'→3')
SB896	γ, F	<aaxab>	1:1	CTGATTACAAGATGAAGGTGG	CAAATTGTATGTATGTGCGGT
SB901	γ	<abxaa>	1:1	CATTGGGTTTTTATCCTATCG	TTAGGGTTTTACACAGAGGAG
SB913	γ, F	<abxab>	1:2:1	TGCTTGTGTAATCTCATGCTC	AGTGAAGGCCTCTCTACCTTT
SB914	γ, F	<abxab>	1:2:1	CTGACCTTCACGCATTCACAC	GATCCTGGTGATCTTGACTTG
SB918	γ, F	<aaxab>	1:1	CCAACTTCGATTCTTGGTCTC	ATGGTGGGCAAAGGAGGTTTA
SB921	γ,	<abxaa>	1:1	CAAAGAAAGACAAGAAAGAGC	ACAGGTGATGATGATAAATCC
SB945	γ, F	<aaxab>	1:1	TACGCCAACAATCTCTCTTAC	GGGCAGTAGAAACTTACAAGG
SB955	γ, F	<abxaa>	1:1	ACCACTCTCCAAATCCCTTAC	ATATTTTAACAAGCCACGCTC
SB984	γ, F	<abxaa>	1:1	ACAATCACACTTCGCATATCA	GGATGGAAAGATTCAAGGATT
SB988	F	<aaxab>	1:1	AAGAAAAGGAGAGAGACCACG	AGAAGAAATGGAGTTGGAGTG
SB1035	F	<aaxab>	1:1	GCAATTACTCAGCTCCTTTTA	TTAGTGTGTTACGTCATTCGA
SB1045	F	<abxaa>	1:1	GTTATCCAAATGTAAGCGAGG	CCTTAGACTGAAATACACCCA
SB1048	F	<abxaa>	1:1	AGTGCTTGTCATTTTCCATAC	ATGTTCTTATTCCCCCTCTTC
SB1060	F	<aaxab>	1:1	AAACACAAGATGAAGAGGAGA	TTAGTTGGAATAGGATGATGG
SB1075	F	<abxab>	1:2:1	AGTATTTCTACCCCCTCTCCC	CCAAGAACATGAAGAGCTACG
SB1084	F	<abxab>	1:2:1	AAAGCAGACACAATGAACACA	TGTGGATTGATGGAAAACTTG
SB1091	F	<abxac>	1:1:1:1	GGTATAGAGTTGGATTGCCTT	GCGTATTTTGTTATCCTGAGA
SB1092	F	<abxab>	1:2:1	AACTGAAACAGGAGCAACATC	AGAGATGGGATACAGGAGAAG
SB1094	F	<abxaa>	1:1	ACTCTATAATTCTTGGCGACC	AGAGATGGGATACAGGAGAAG
SB1148	F	<abxaa>	1:1	TAGGATGTTTCTGAGGCTTTC	AGACTTGCTAGAGACTTGGCC
SB1185	F	<aaxab>	1:1	TGAGGTCATGGTTGAGTTATG	ATGGTGCCTGCAATCTTTAAC
SB1196	F	<abxab>	1:2:1	ATTCTCCGTCACCATAACCAC	GATGCTATCGAAGGGATGAAC
SB1229	F	<aaxab>	1:1	AAGCTAACCGGAACAATACCA	TCAATAAAGAAAGGGAGACCA
SB1249	F	<abxaa>	1:1	GTGCTCAAATTAAAGGGAGAT	CGCTCGATGAACAAACTCTAC
SB1254	F	<abxab>	1:2:1	GAATAGTTATCGCACTCTTGG	GGTGGTGAAGCCTGTAGTGTA
SB1303	F	<abxaa>	1:1	GCATTTCCAATCTCACAACAA	TGGTCTGCTTATGCTCTTATC
SB1308[d]	F	<abxab>	3:1	CCTTGTTTCGTGACTGATATG	TCACTCATGGCTTTTTTAGAC
SB1317	F	<abxaa>	1:1	AGCATACGAGGTTCTTAGTGA	CCACCAAAGAAGGATTGAAAG
SB1318	F	<aaxab>	1:1	TGATTAGCATGTTCTGAGGTT	TGCTCAGTTCCTTTTAACTGT
SB1332	F	<abxac>	1:1:1:1	CTTCCCCGTCTTTAATTTCAG	TCCAGCATCATAAAAGATTCG
SB1357	F	<abxab>	1:2:1	TATAGCTGTGGAGAGTCGGGT	AGTCCCTAGCAAAATGACCAG
SB1366	γ	<abxab>	1:2:1	TCAATACATCCATCCTAGAGG	ATCCCTTTTCACATTCACCTG
SB1464	γ	<abxab>	1:2:1	TTTTCGCTACTTCATACCAGA	ACCACACAACCAAACAATAAG
SB1518	γ	<abxab>	1:2:1	AATTAGCTTCGGACACATTTC	AACCTCATCATCTTCAACAAC

γ: marker detected by radioisotope methods
F: marker detected by fluorescent methods
Ag: marker resolved by agarose gel electrophoresis
d: indicates marker was scored as a dominant marker

Distorted segregation ratios were observed for ten (14%) of those microsatellites screened ($P < 0.1$).

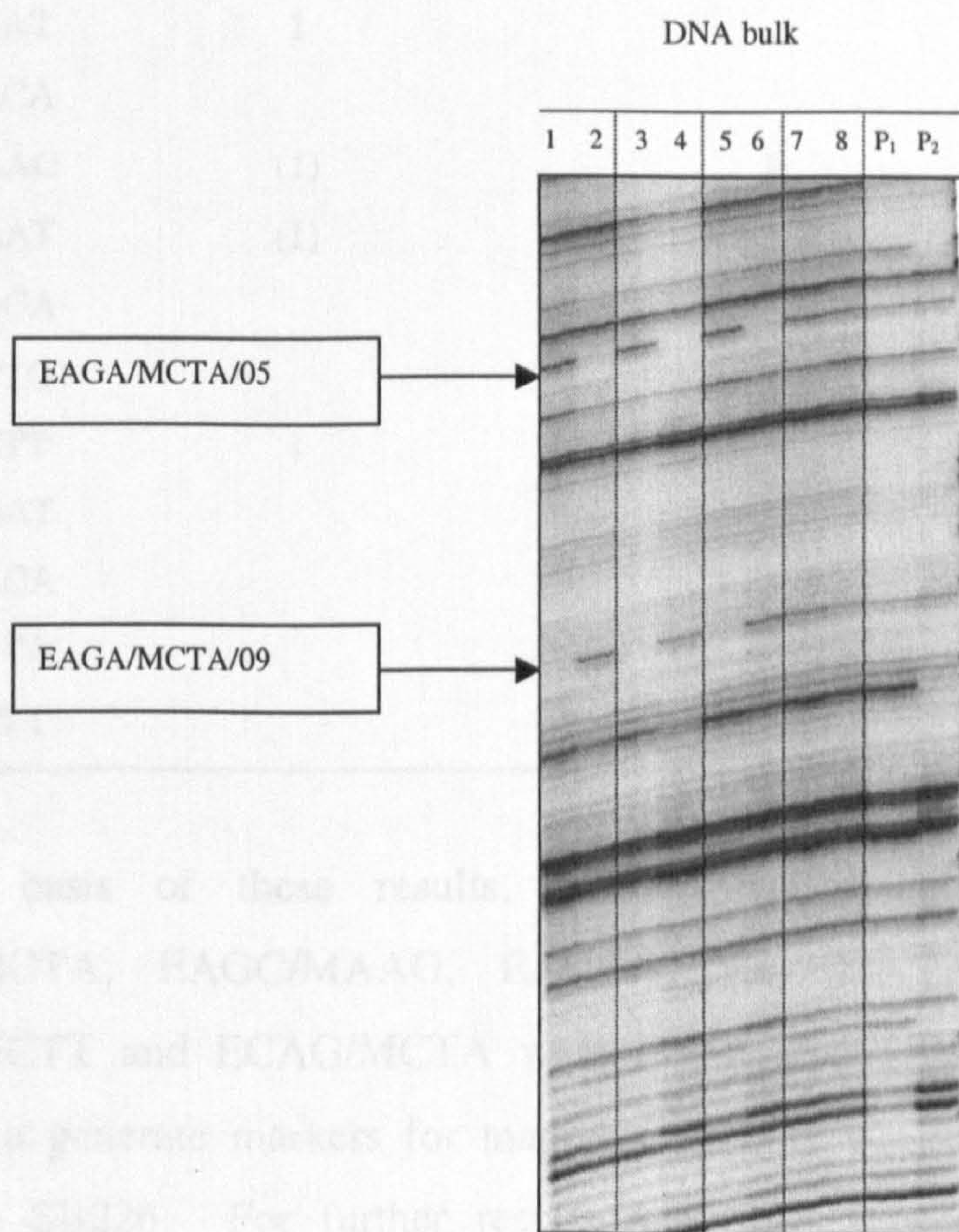
5.3.6 Bulked segregant analysis

AFLP profiles for all bulked DNA samples were obtained for 44 of the 56 primer combinations used, with all markers detected in the parents also being identified in the

bulked samples. The eight primer combinations involving the EAAC primer generated extremely faint profiles that could not be scored. Similar results were obtained for primer combinations EATA/MCCG, EATA/MCTA, EATA/MCTG and EATA/MCTT. For seven potentially interesting markers, accurate scoring was difficult due to loss of resolution towards the bottom of the gel or because they were located extremely close to an adjacent band. From the 44 successful primer combinations, 10 (+5 that were difficult to score) markers were highlighted as putatively linked to microsatellite locus SB226 and 17 (+4) markers were identified that differed between the field-based rust phenotype pools. Of these, 2 (+2) markers were identified that segregated between the alternative SB226-targeted bulks and also the rust resistant and susceptible bulks.

An example of a bulked segregant profile generated with primer combination EAGA/MCTA indicating two AFLP markers putatively linked to locus SB226 is shown in Figure 5.5.

Figure 5.5. Bulked segregant analysis AFLP profiles identifying two markers putatively linked to microsatellite locus SB226.



The results of BSA for those primer combinations yielding putatively informative markers are summarised in Table 5.5.

Table 5.5. Details of primer combinations yielding putatively informative markers from BSA. The number of markers that segregated only between the SB226-targeted bulks, only between the rust resistant/susceptible bulks and between both the SB226-targeted and rust resistant/susceptible bulks are shown. Figures *in brackets* indicate a marker that could not be conclusively scored.

Primer combination	Markers putatively linked to locus SB226	Markers putatively linked to rust resistance/susceptibility	Markers linked to both
EAGA/MAAG	2		
EAGA/MCTA	2		
EAGA/MCTT		(1)	
EAGC/MAAG	(1)	2	1
EAGC/MACC		1	
EAGC/MCTA	1	1	
EAGC/MCTT		2	
EACT/MACC		1	
EACA/MACC		1	
EACA/MACA		1	
EATA/MAAG	1	1	
EATA/MAAT	1		
EATA/MACA			(1)
ECAA/MAAG	(1)	1	
ECAA/MAAT	(1)		
ECAA/MACA			(1)
ECAA/MCTG		1	
ECAA/MCTT	1	1	
ECAG/MAAT		(1)	
ECAG/MACA		1	
ECAG/MCTA			1
ECAG/MCTT		1	

On the basis of these results, AFLP primer combinations EAGA/MAAG, EAGA/MCTA, EAGC/MAAG, EAGC/MCTA, MATA/MAAG, EATA/MAAT, ECAA/MCTT and ECAG/MCTA were highlighted as those that should be used initially to generate markers for mapping, as they would potentially yield markers linked to SB226. For further results regarding mapping of the AFLP markers highlighted by the BSA approach see Section 5.3.8. Results obtained from the resistant versus susceptible bulks were not used as a basis for selection of informative

AFLP primer combinations as only a one resistant and one susceptible pooled DNA sample were included in the experiment. As individuals comprising these bulks were not screened individually for the presence or absence of putatively informative markers, there was a strong possibility that several of the highlighted markers may have been false positives.

In addition to the primer combinations selected from screening the bulks, the primer combinations EAGC/MAAT and EAGC/MACA were included for mapping as they gave parental profiles that included a relatively large number of polymorphic markers.

5.3.7 AFLP markers

In total, 149 polymorphic AFLP markers were scored from seven primer combinations, the characteristics of which are shown in Table 5.6.

Table 5.6. Characteristics of polymorphic AFLP markers generated from seven primer combinations. Restriction enzymes/adapters/primers used in AFLP production are indicated for each combination (E = *Eco*RI, M = *Mse*I and P = *Pst*I). Additional characters indicate selective nucleotides.

Primer combination	Female markers[a]	Male markers[b]	Heterozygous markers[c]	Total polymorphic markers
EAGA/MAAG	6	4	14	24
EAGA/MCTA	3	5	8	16
EAGC/MAAG	4	6	15	25
EAGC/MAAT	7	9	6	22
EAGC/MACA	3	2	9	14
EATA/MAAG	8	8	7	23
EATA/MAAT	7	4	14	25
Total	**38**	**38**	**73**	**149**

a: *abxaa* marker; present in the female parent only; segregating 1:1 in the progeny
b: *aaxab* marker; present in the male parent only; segregating 1:1 in the progeny
c: *a0xa0* marker; heterozygous in both parents; segregating 3:1 in the progeny

The average total number of bands identified per assay was 67, with an average of 20 (29.9%) revealing scorable polymorphisms in the K8 population. Following analysis of genotype frequencies for each AFLP marker, 17 showed segregation distortion ($P <$ 0.1: chi-square test). Thirteen highly distorted markers with P values less than 0.0005 (~8.7% of markers) were removed from the data set prior to construction of the

parental and consensus linkage maps. Of these, four were of segregation type *abxaa*, four were of type *aaxab* and five were *a0xa0* markers.

5.3.8 Construction of the parental maps

For the paternal map, a total of 34 AFLP markers and 28 microsatellite markers that segregated in the mapping population were included in linkage analysis. At a LOD threshold of 4.0, 43 (69.4%) markers could be assigned to eight linkage groups each containing a minimum of three markers. In addition, five duplets were formed and nine markers remained unlinked. For the majority of groups, map construction proved straightforward with no markers causing problems with respect to goodness of fit. However, two markers were removed from the data set as they were problematic in map calculation. The first, EATA/MAAT/1 could not be easily placed on to the Group A map. This marker was difficult to score and yielded a distorted segregation ratio ($P < 0.01$), suggesting that genotype data for this locus may have been erroneous. The second difficult marker, EATA/MAAG/17, showed linkage to group P. However, when included in subsequent analysis, this marker prevented calculation of the linkage group. This marker had been noted as difficult to score upon initial examination of the autoradiograph and was, therefore, discarded. Calculation of marker orders and relative map distances could not be achieved for Group D when only marker pairs with LOD significances greater than 1.0 were included in analysis. However, when this threshold was relaxed to incorporate marker pairs with LOD significances greater than 0.9, all markers assigned to this group were mapped without difficulty.

The paternal map comprised 28 AFLP markers and 22 microsatellites. The resulting 13 linkage groups (including duplets) spanned 490.0 cM with an average interval between markers of 13.2 cM. (Figure 5.6).

For the maternal map, a total of 34 AFLP markers and 30 microsatellite markers were included in linkage analysis. At a LOD threshold of 4.0, 53 (82.8%) markers could be assigned to nine linkage groups each containing a minimum of three markers. In addition, three duplets were formed and 11 markers remained unlinked. Two AFLP markers, EAGA/MAAG/22 (Group T) and EAGA/MAAG/5 (Group F) were not mapped as they could not be positioned without using the capability of JoinMap

software to force markers on to the linkage map. Calculation of linkage maps for all other groups proved unproblematic.

The maternal map comprised 27 AFLP markers and 24 microsatellites. The resulting 12 linkage groups (including duplets) spanned 528.7 cM with and average interval between markers of 13.6 cM. (Figure 5.7).

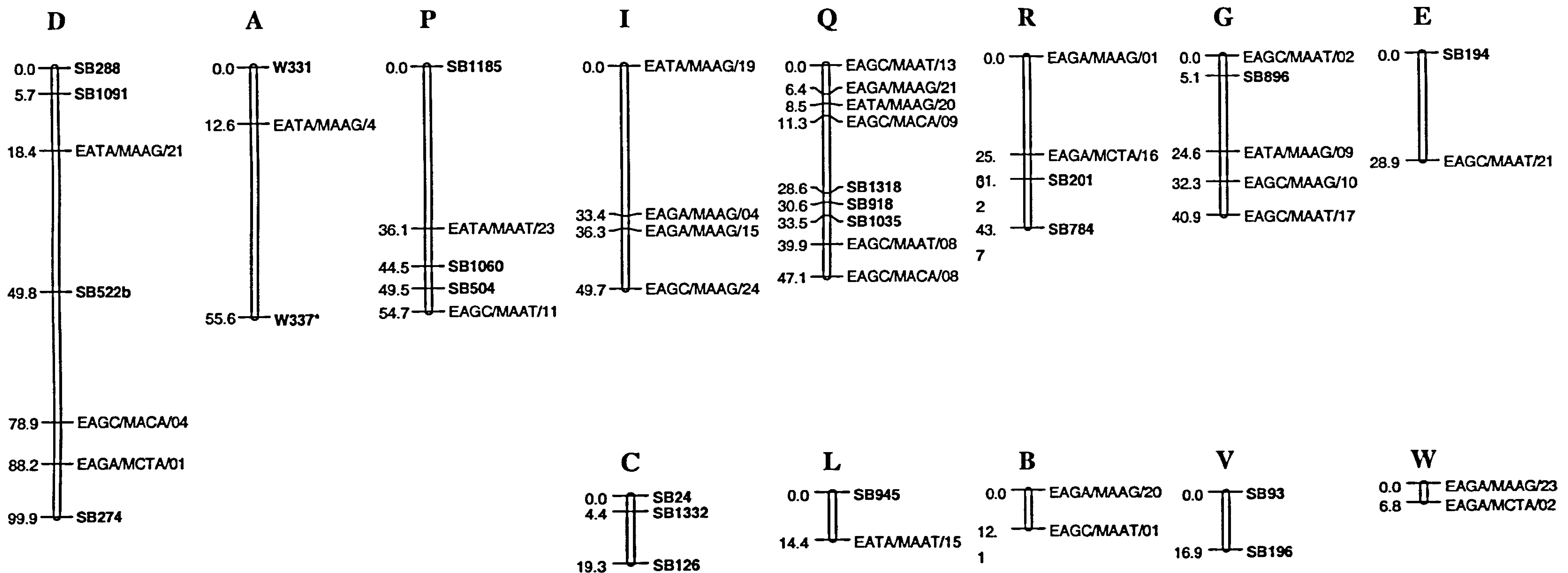


Figure 5.6. The K8 paternal linkage map. Linkage groups are presented in length order with group names assigned in accordance with K8 consensus group nomenclature. Marker names are shown on the *right* of each group with map distances (in cM) indicated on the *left*. For details of AFLP marker nomenclature see Table 5.6. Microsatellite markers are indicated in *bold type* with the prefix SB-. Markers showing significant levels of segregation distortion are indicated by *asterisks* (*: $P < 0.1$).

Figure 5.7. The K8 maternal linkage map. Linkage groups are presented in length order with group names assigned in accordance with K8 consensus group nomenclature. Marker names are shown on the *right* of each group with map distances (in cM) indicated on the *left*. For details of AFLP marker nomenclature see Table 5.6. Microsatellite markers are indicated in ***bold type*** with the prefix SB-. Markers showing significant levels of segregation distortion are indicated by *asterisks* (*: $P < 0.1$, **: $P < 0.05$, ***: $P < 0.01$, ****: $P < 0.005$).

5.3.9 The K8 consensus map

For the consensus map, those markers identified as problematic whilst calculating the parental maps were excluded from linkage analysis. The final mapping data set comprised 34 *abxaa* AFLP markers, 34 *aaxab* AFLP markers, 69 *a0xa0* AFLP markers, and the 71 microsatellite markers shown in Table 5.4. Marker groupings at a LOD threshold of 4.0 resulted in the formation of 24 linkage groups comprising three or more markers, with two duplets and 12 markers remained unlinked. With the exception of three groups, marker orders and relative map positions were calculated without difficulty.

For the first problematic group, a map could not be calculated due to a potentially spurious linkage of two groups (H and I) caused by the presence of microsatellite marker SB587 (type *abxac*). No evidence of linkage was observed between these groups in the paternal map, which also included segregation data for paternal alleles from SB587. Furthermore, this locus remained unlinked to any other markers in both paternal and maternal linkage analyses, even at a decreased LOD threshold of 3.0. In response, this marker was removed from the data set and marker groupings were recalculated, resulting in the identification of 25 major linkage groups, with the number of duplets and unlinked markers remaining unchanged. For group I, inclusion of an *a0xa0*-type AFLP marker (EATA/MAAG/14), previously highlighted by BSA as being linked to microsatellite locus SB226, caused a disagreement between marker orders of the corresponding paternal and consensus groups. This marker was difficult to score on a number of autoradiographs and was, therefore, discarded from further analysis. Subsequent recalculation of this consensus linkage group resulted in congruent marker orders between maps.

A discrepancy between marker orders of the maternal and consensus maps was also observed for the second problematic linkage group, Group A. Here, the inclusion of a single *a0xa0* AFLP marker was shown to be the cause. When excluded from analysis, resulting marker orders were in agreement between maps.

For group Y, which comprised only markers segregating in both parents, three AFLP markers of type *a0xa0* were excluded from map construction as they were shown to cause 'tension' within the group with regard to goodness-of-fit. If included, all markers assigned to this group could not be placed without using the capability of the

software to force them on to the map. When removed from this linkage group, remaining markers could be placed without difficulty.

A total of 18 consensus linkage groups (A – O and X - Z), defined here as those containing one or more markers that segregated in both parents, could be calculated. Of these, three groups (X, Y and Z) did not contain any markers that segregated exclusively in either parent. The consensus linkage groups[†] comprised 19 *abxaa* AFLP markers, 20 *aaxab* AFLP markers, 49 *a0xa0* AFLP markers and 50 microsatellites. The total map length (including duplets) of these groups was 1191.8 cM with an average interval between markers of 10.1 cM (Figure 5.8).

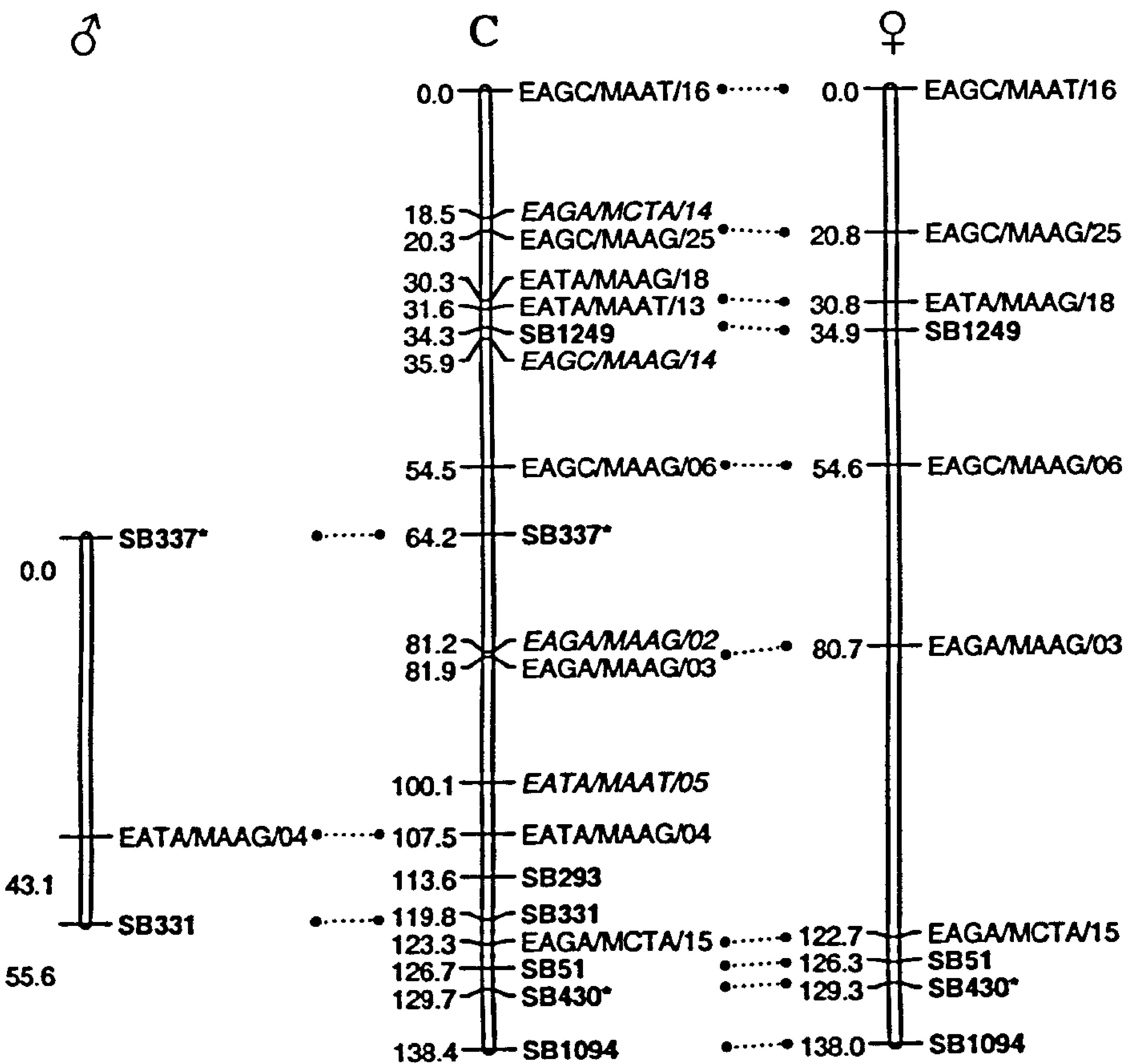
Linkage groups P – W comprised only markers segregating in a single parent, hence consensus linkage groups could not be calculated in these instances. Groups P, Q, R, V and W contained paternal markers exclusively and represented a combined map length of 169.2 cM. Groups S, T and U comprised maternal markers only and spanned a total map distance of 112 cM.

Both AFLP and microsatellite markers were well distributed over the linkage groups identified, although some evidence of AFLP clustering was observed on consensus linkage groups A and B.

5.3.10 Mapping AFLP markers putatively linked to microsatellite SB226

Of the eight AFLP primer combinations identified by BSA as potentially yielding markers linked to microsatellite locus SB226, five were used to generate markers for mapping. Of the seven markers identified by BSA as being putatively linked to SB226, six were successfully mapped to the target region on linkage group I (Figure 5.8). These markers mapped to within 17.3 cM of locus SB226. Despite evidence of strong linkage to group I (linked at a LOD threshold of 10.0), the seventh marker, EATA/MAAG/14, was not mapped for the reason described in Section 5.3.9.

[†] linkage groups comprising at least one marker heterozygous in both parents



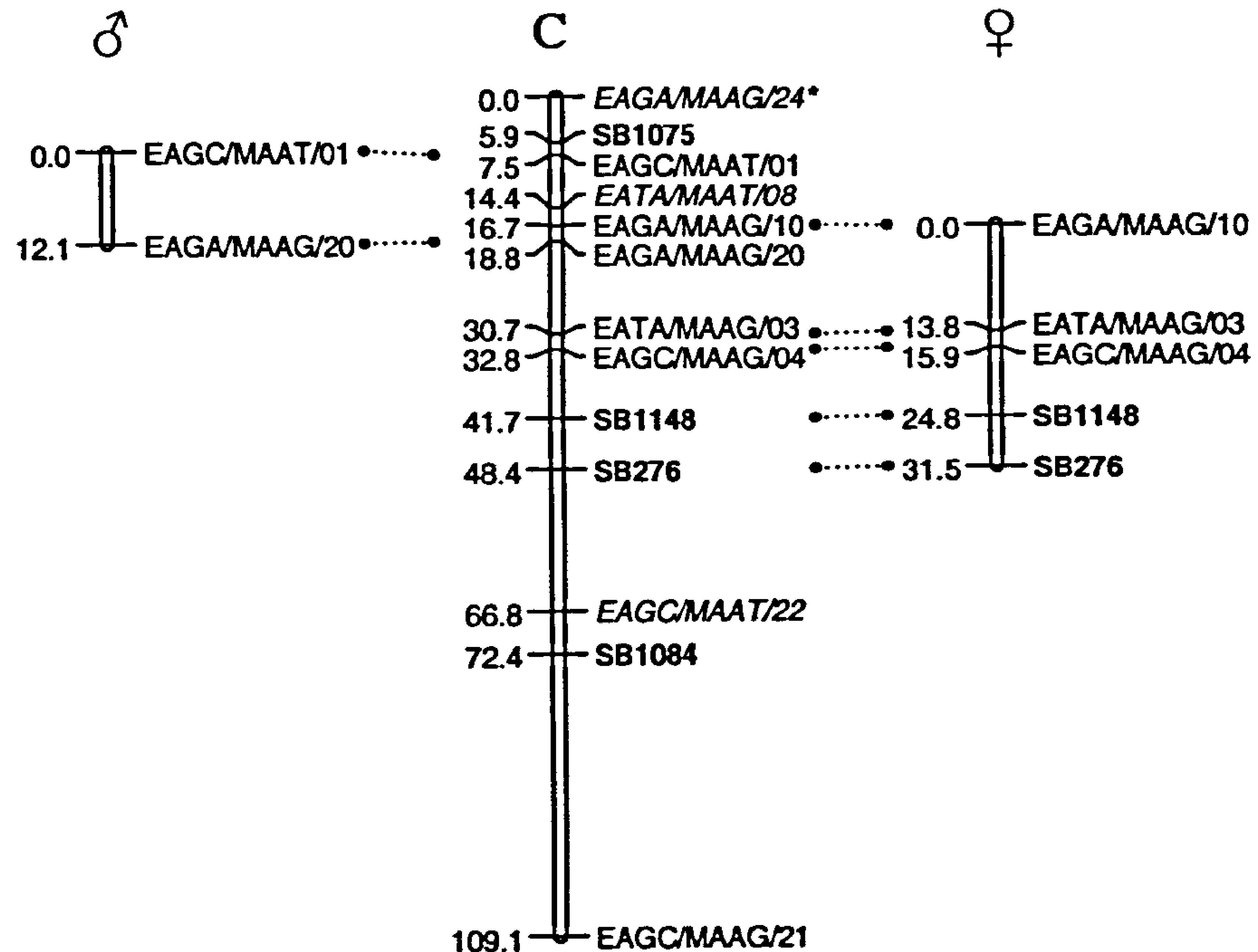


Figure 5.8. The K8 consensus map. Alignments between paternal (♂), maternal (♀) and consensus (C) linkage groups are shown. Corresponding K3 linkage groups (Figure 5.9) are indicated *in superscript* text next to K8 group names. Marker names are shown on the *right* of each group with map distances (in cM) indicated on the *left*. Microsatellite markers are indicated in **bold type** with prefix SB-. For details of AFLP marker nomenclature see Table 3.2. AFLP markers of type *a0xa0* are shown in *italics*. Corresponding locus positions are indicated by *dotted bars* between maps. Markers showing significant levels of segregation distortion are indicated by *asterisks* (*: $P < 0.1$, **: $P < 0.05$, ***: $P < 0.01$, ****: $P < 0.005$, *****: $P < 0.001$). Marker identified as putatively linked to microsatellite locus SB226 by bulked segregant analysis are indicated by the letters *BSA*.

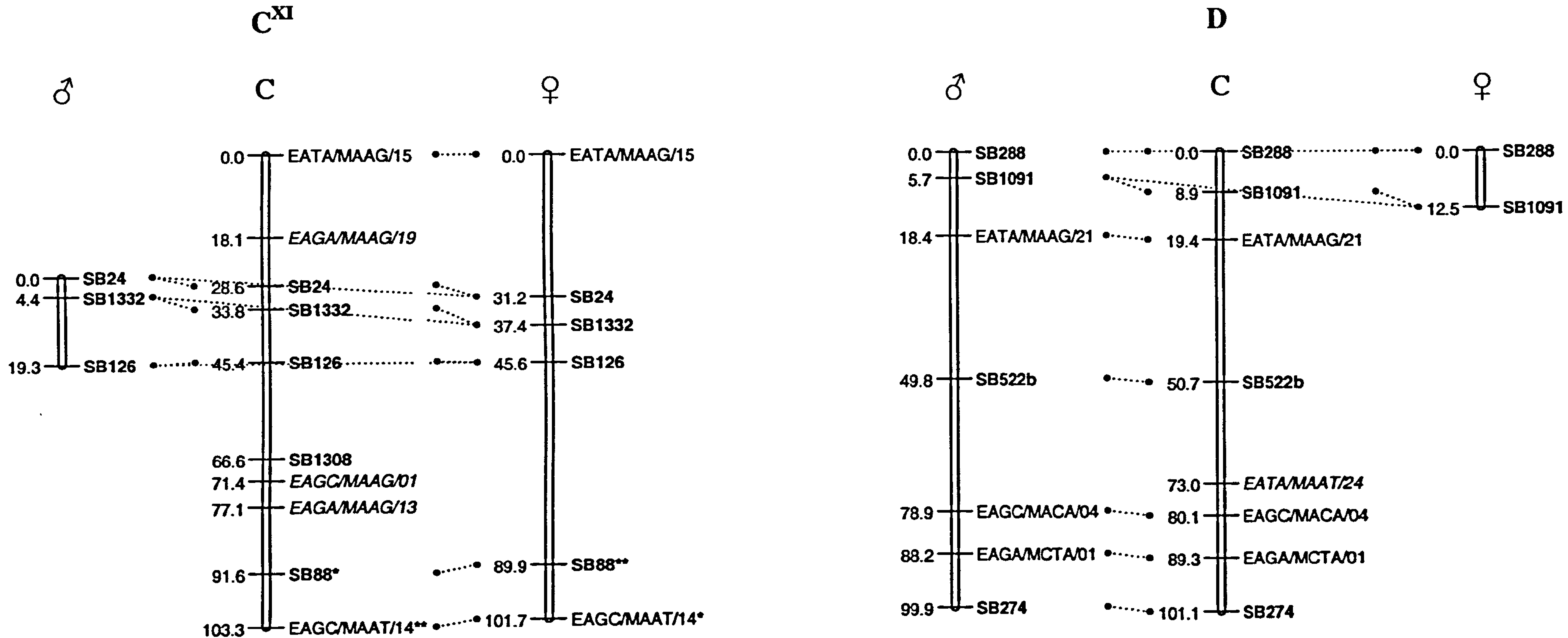
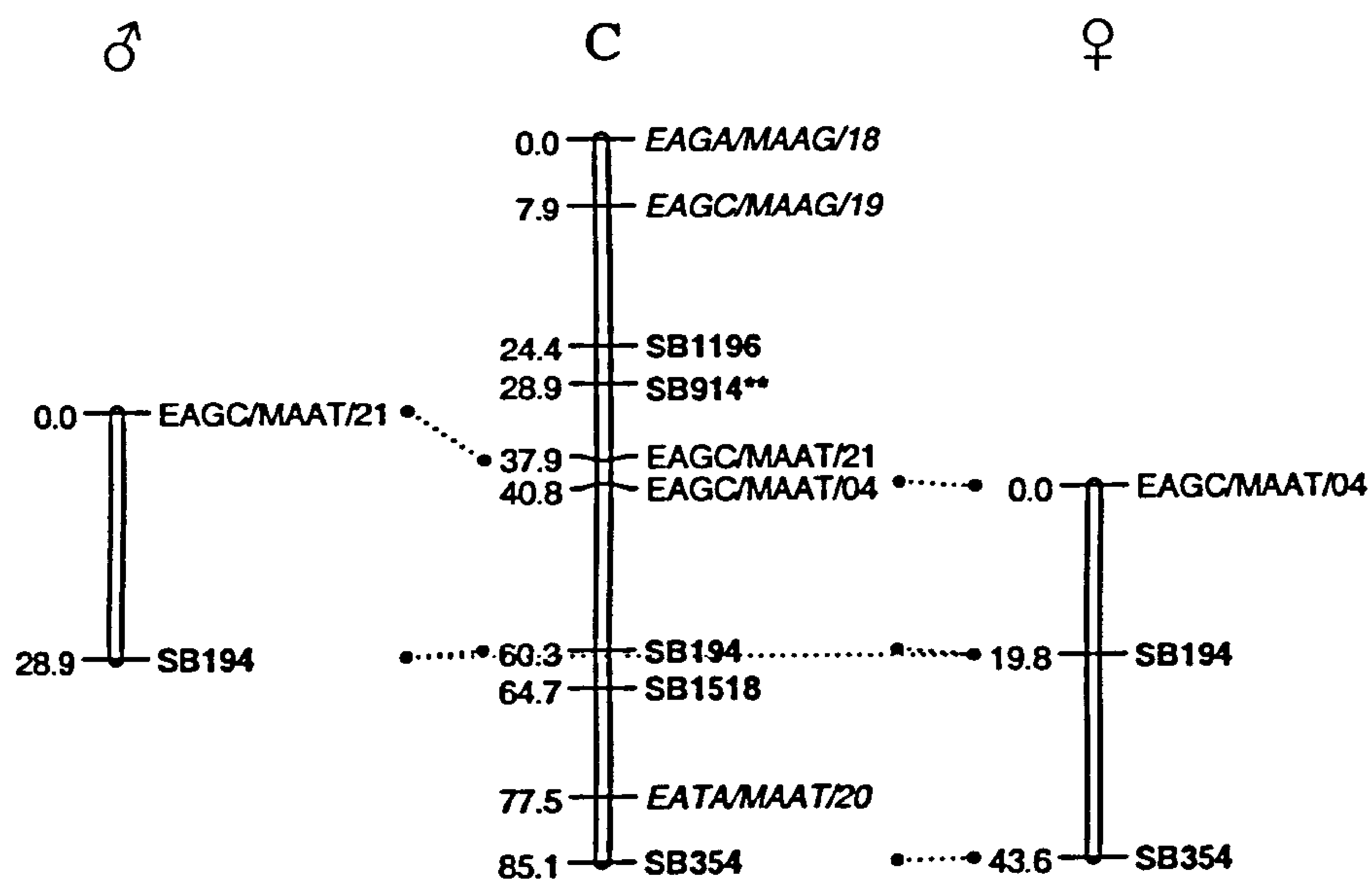


Figure 5.8 continued. The K8 consensus map.



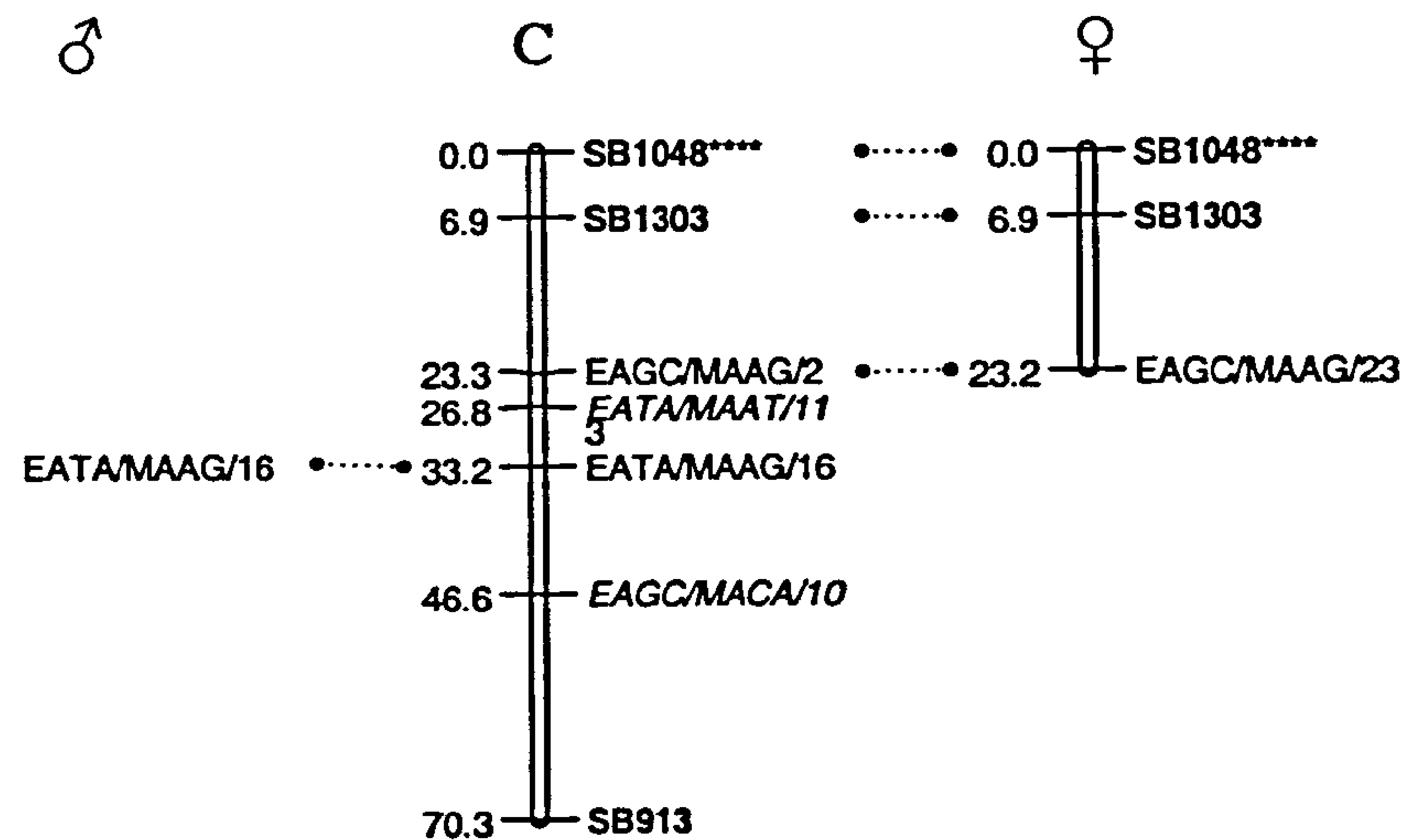


Figure 5.8 continued. The K8 consensus map.

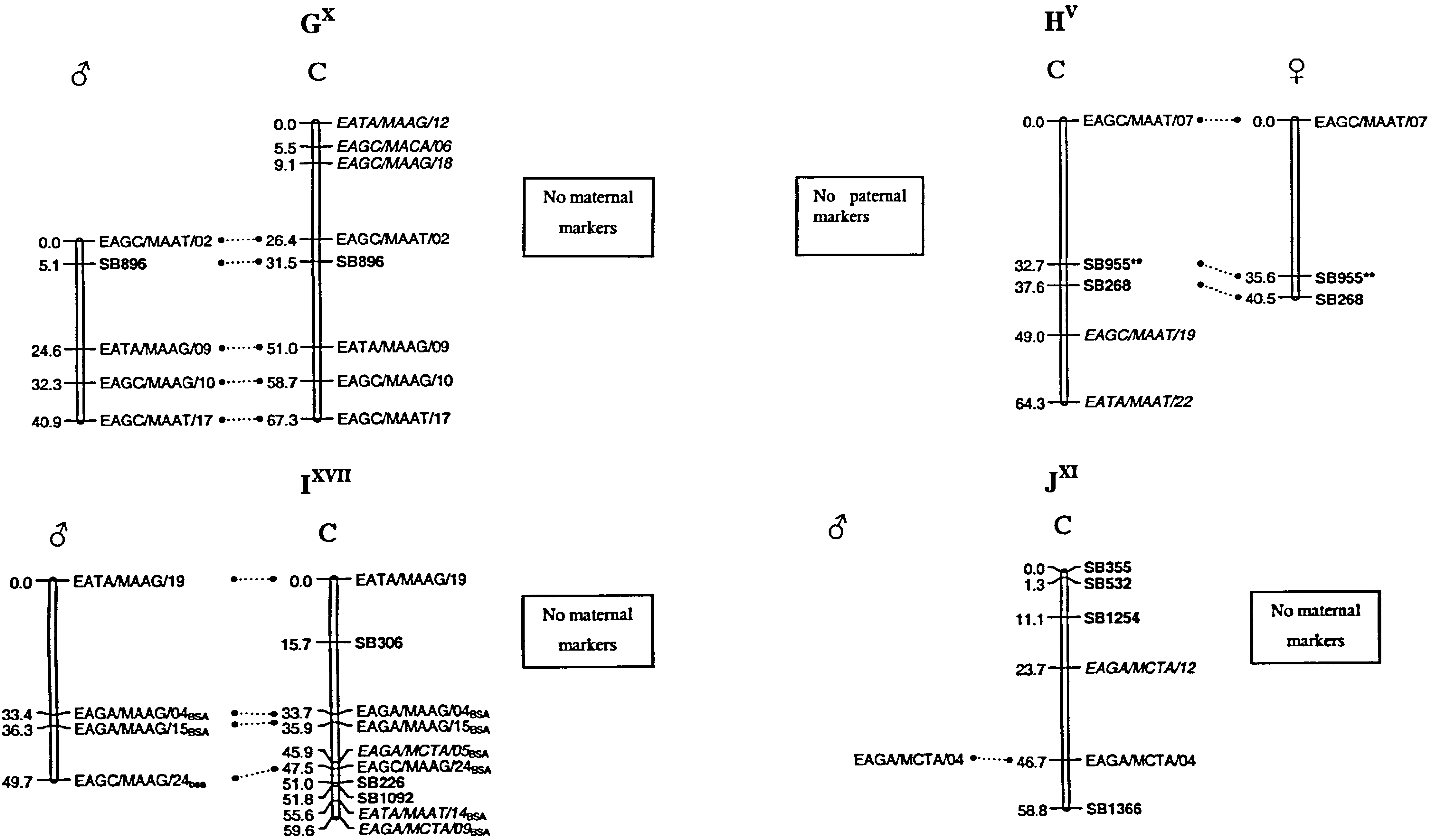
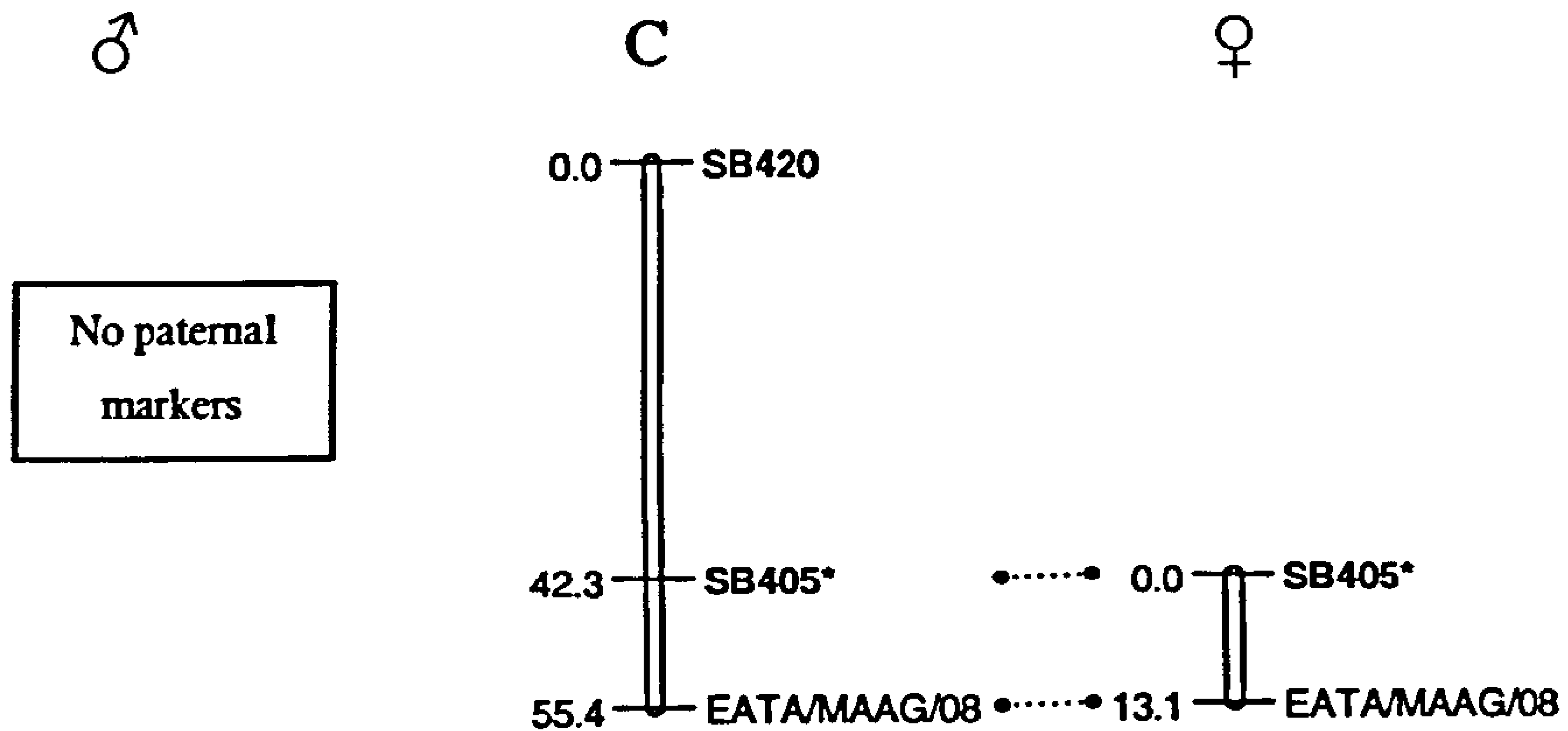
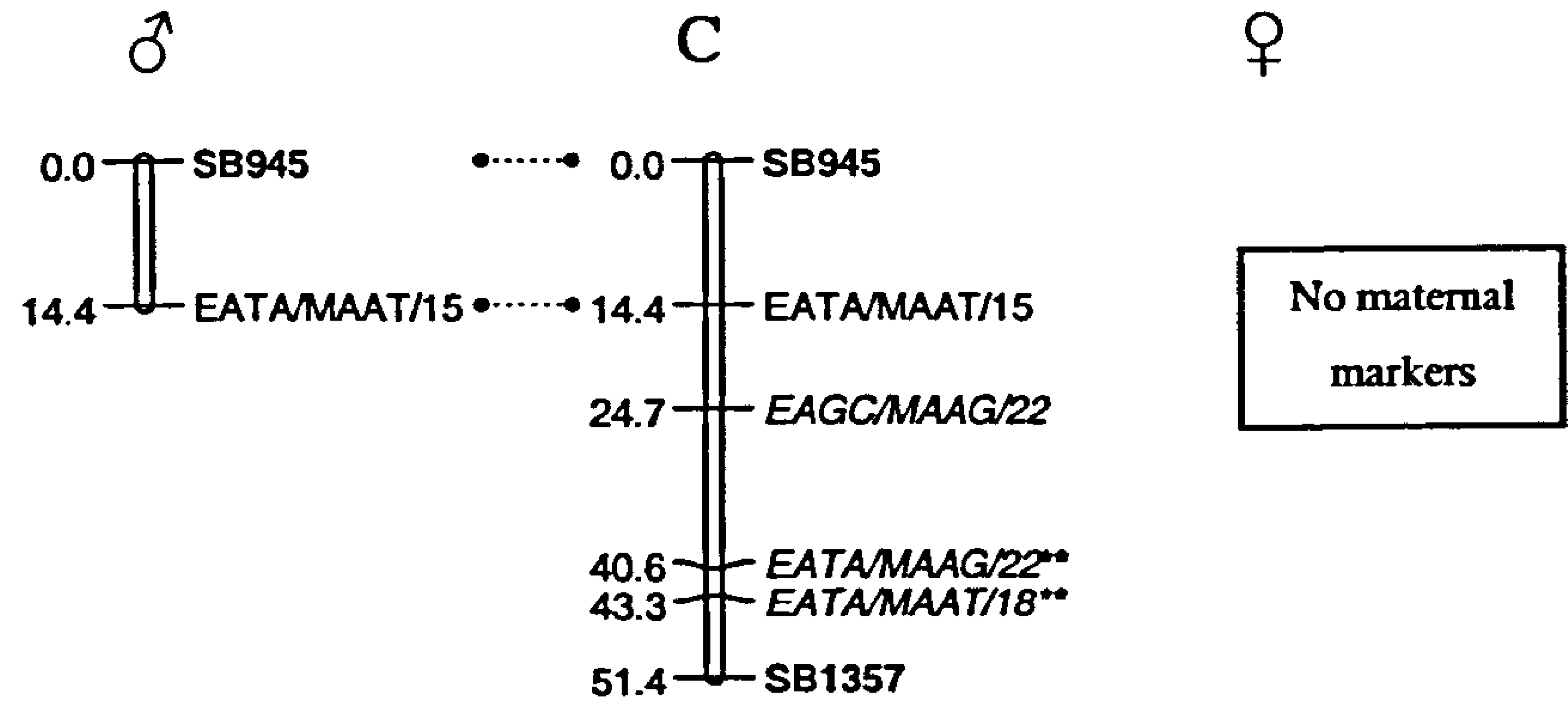
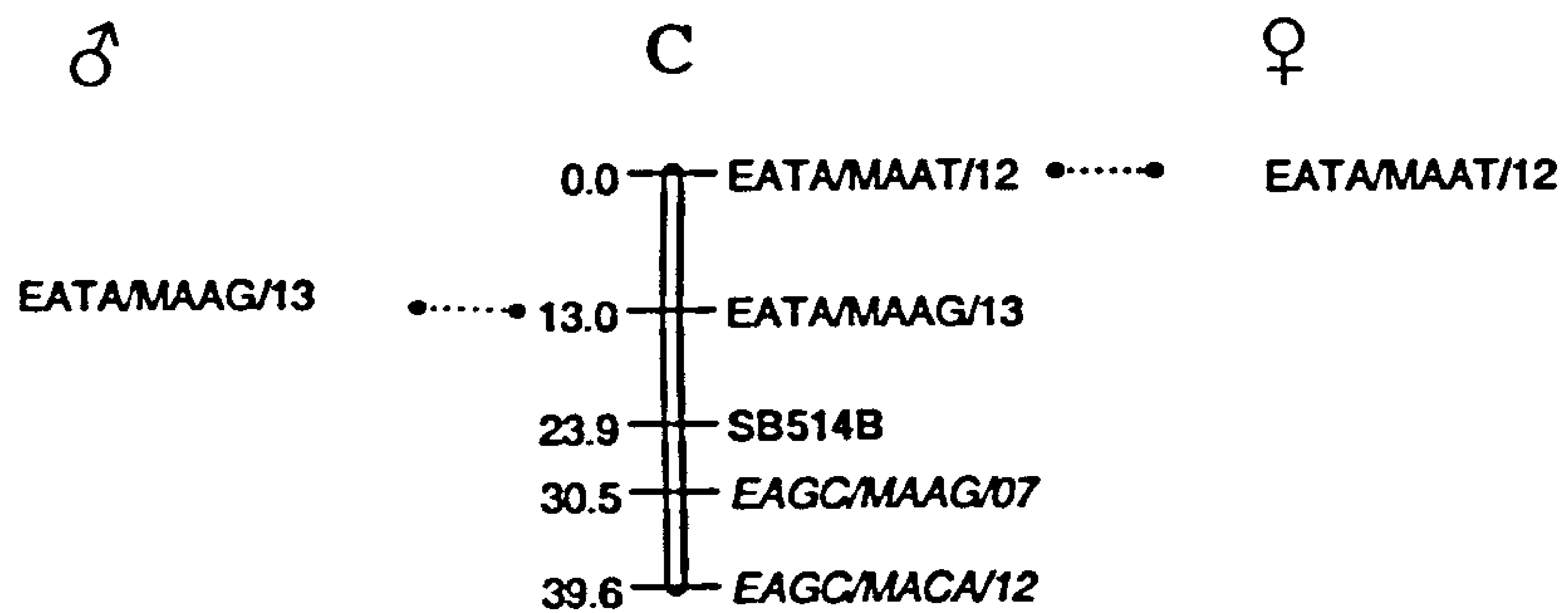


Figure 5.8 continued. The K8 consensus map.







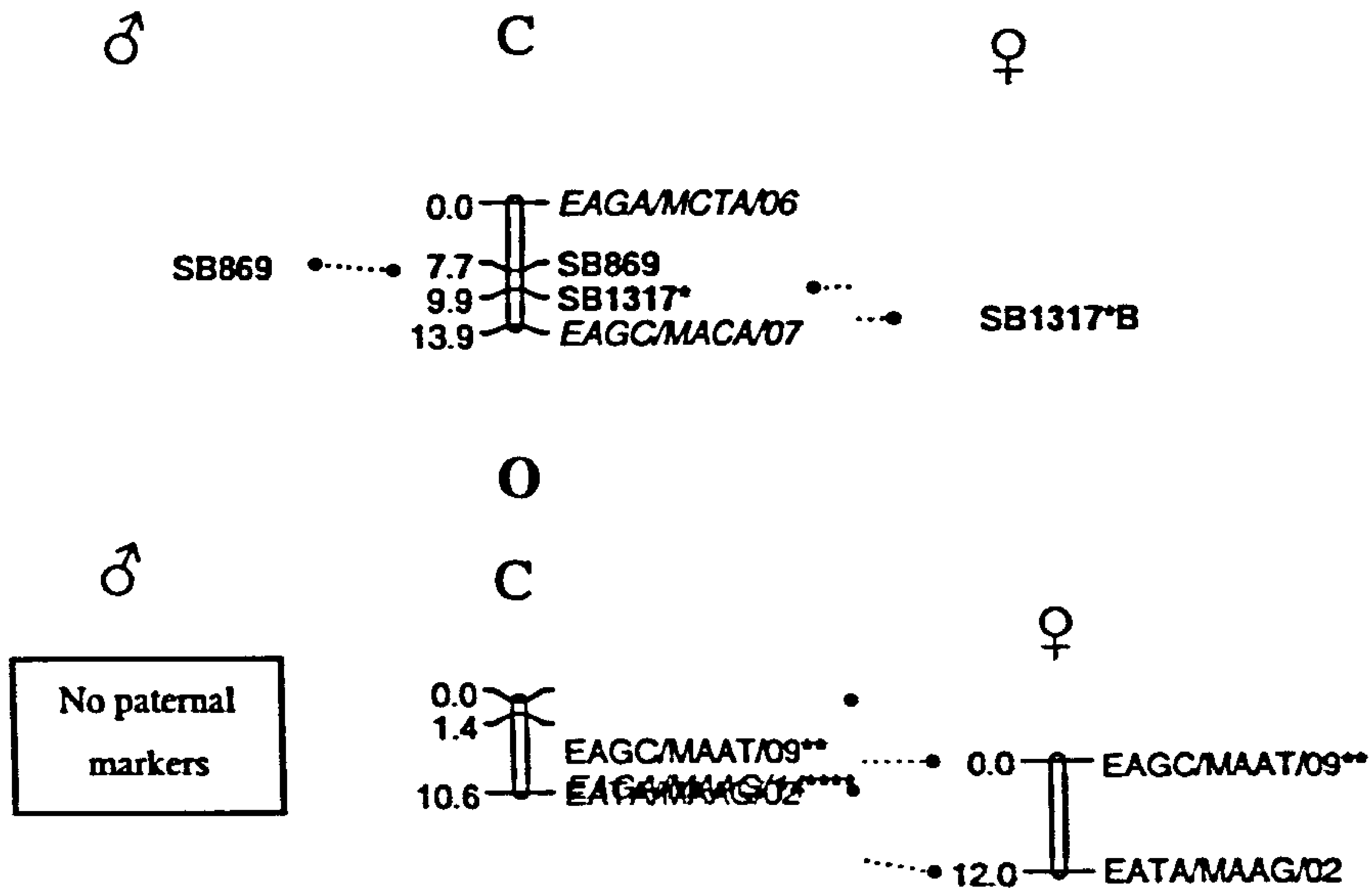


Figure 5.8 continued. The K8 consensus map.

Figure 5.8 continued. The K8 consensus map.

5.3.11 Identification of homeologous K3 and K8 linkage groups

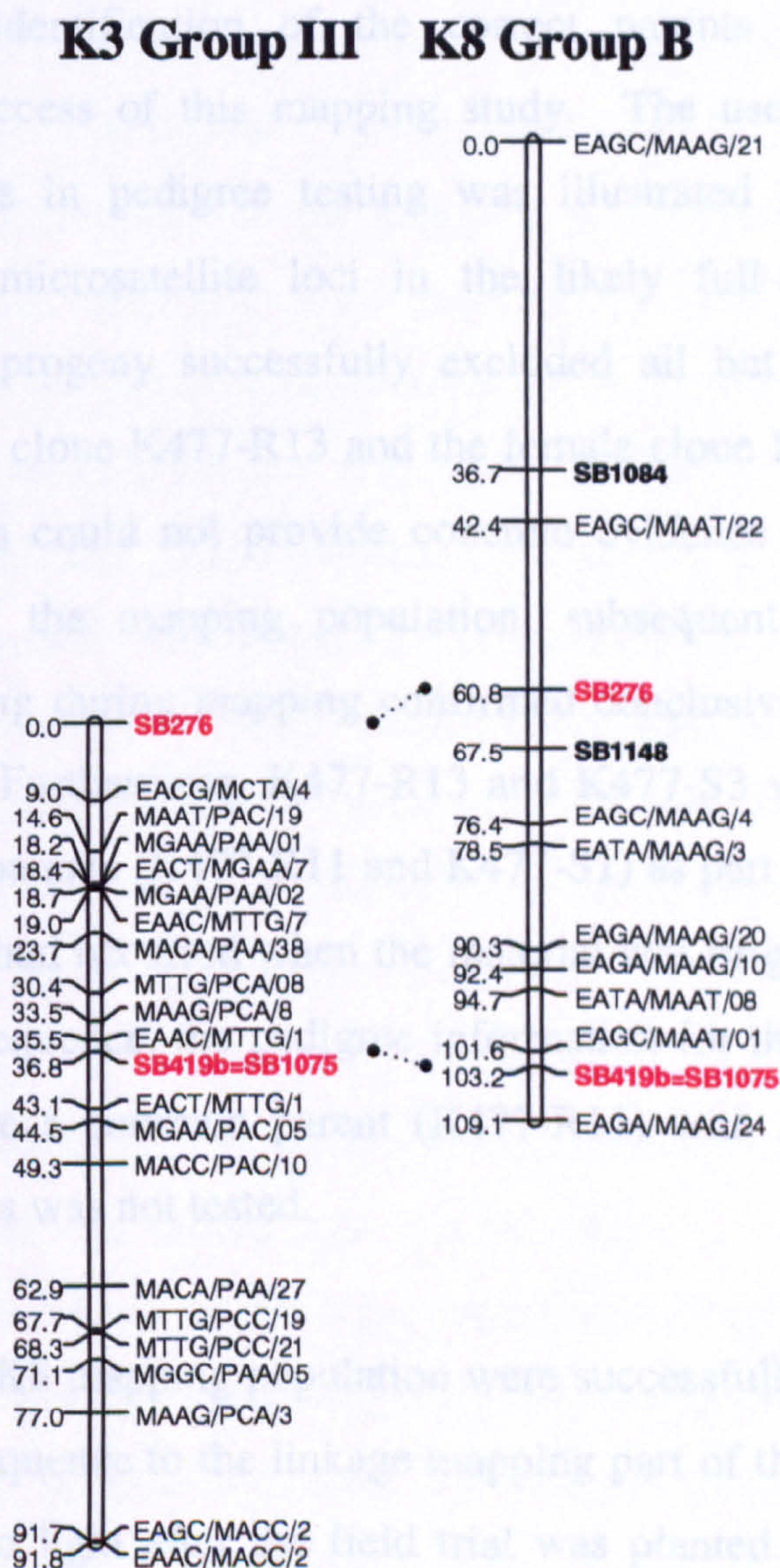
Microsatellite markers mapped in both the K3 and K8 populations were used as locus-specific anchor points between the two respective consensus maps. For the most part, the relative positions of these 25 microsatellites were in agreement across both maps, with no sets of markers showing linkage in one map but not the other. Putative identification of 17 homeologous linkage groups was achieved, although the orientation of corresponding linkage groups relative to one another could not be deduced for groups aligned by a single microsatellite marker (Figure 5.9). However, one discrepancy was observed. In K3, microsatellites SB274 and SB504 both mapped to linkage group IV, while in K8, these markers mapped to linkage groups D and P, respectively. Although group P comprised only paternal markers, group D contained four markers heterozygous in both parents. Hence, if these groups were truly linked in K8, detection of linkage would have been likely given the marker information available. These results suggested that in K8, these primers did not detect linked polymorphisms.

Figure 5.9. Alignment of putatively homeologous K3 and K8 linkage groups. Microsatellite markers used to anchor corresponding linkage groups are shown *in red type.* K3 linkage groups are indicated by *roman numerals* unless otherwise stated. K8 linkage groups are indicated by standard *alphabetic* names.

Figure 5.9 continued.

In K8, microsatellite markers SB276 and SB1075 both mapped to linkage group B. Locus SB1075 is homologous to locus SB419b, for which, alternative primers were designed as the original set gave a complex banding pattern when tested on K8 parental DNA. Although SB419b was not mapped on K3 for reasons discussed in Section 3.3.5, evidence of linkage to group III was observed. However, if the ability of JoinMap software to force this marker on to the map was used (an approach not routinely employed in K3 map construction) SB419b was found to map to a position 37 cM away from SB276 (Figure 5.10). This result further substantiates the claim that groups III and B of the K3 and K8 linkage maps, respectively, represent homeologous linkage groups.

Figure 5.10. Alignment of K3 linkage group III and K8 linkage group B based on microsatellite loci SB276 and SB419b (=SB1075).



5.4 Discussion

In this chapter, the construction of a framework genetic linkage map for K8 is described. Microsatellite markers enable the anchoring of this map to the K3 map. The fact that good co-alignment was found between the maps confirms the robustness of the linkages established and provides a firm foundation for subsequent QTL analysis (Chapter 6). Furthermore, the availability of two genetic linkage maps of willow (K3 and K8), which share several locus-specific microsatellite markers provides the means for future comparative mapping studies, not only in terms of linkage analysis, but also for comparative QTL analysis.

5.4.1 Identification of the K8 parents

Following the discovery that the pedigree information supplied for the K8 population was incorrect, the identification of the correct parents became of paramount importance to the success of this mapping study. The usefulness of co-dominant microsatellite markers in pedigree testing was illustrated here. Interrogation of segregation at five microsatellite loci in the likely full-sib parents (the K477 population) and K8 progeny successfully excluded all but two possible potential candidates – the male clone K477-R13 and the female clone K477-S3. Although this microsatellite analysis could not provide concrete evidence that these two willows were the parents of the mapping population, subsequent AFLP and additional microsatellite screening during mapping confirmed conclusively that the true parents had been identified. Furthermore, K477-R13 and K477-S3 were planted in adjacent plots to the expected parents (K477-R11 and K477-S1) as part of the NWC, suggesting that a sampling error had occurred when the material was originally collected to make the cross. As a consequence, the pedigree information for the K1 population, which was expected to share a common parent (K477-R11) with K8, is also likely to be incorrect, although this was not tested.

As the parents of the K8 mapping population were successfully identified, the original error was of no consequence to the linkage mapping part of this study. However, this problem only came to light after the field trial was planted. Therefore, the correct parents were not included in the trial and were not included in the phenotypic

assessments. Furthermore, they were also not included in the laboratory-based trait assessments of rust and beetle palatability performed during the establishment year. While subsequent detection of QTL will not be directly affected by this problem, more fundamental trait analyses, such as calculation of heritabilities will not be possible with the current data set.

5.4.2 Microsatellite markers

5.4.2.1 Microsatellite marker development

The need to develop additional microsatellite markers for mapping in K8 arose from the observation that a large proportion of previously developed markers (Section 3.3.2) either failed to detect polymorphism in the K8 mapping population or failed to generate a profile that could be reliably scored. Difficulties in detecting polymorphic microsatellite loci were not unexpected given the K8 pedigree, i.e. in this pedigree *S. viminalis* var. Astrid, a great grandparent of the mapping population, was backcrossed to one of its offspring, "SW930984". From the full-sib progeny of this cross, two individuals were then crossed to produce the K8 population. This pedigree construction would have resulted in lower levels of polymorphism than a wide cross between two unrelated parents. The observation that several microsatellite primer sets yielded profiles that were easily scored in K3, but were difficult to interpret in K8, may be a result of the inclusion of *S. schwerinii* in the background of K8 pedigree. The K3 pedigree comprised only *S. viminalis* and was used for initial testing of the original microsatellite primer sets. This result also suggested that, although not tested on K8 due to lack of time during the current study, some microsatellites that were not useful in the K3 population due to complex profiles or failure to yield a PCR product, may be informative in K8 and could be pursued in future studies.

The re-design of microsatellite primers sets to loci that were not successfully amplified previously using the original primer sets, was beneficial in some cases. For example, primer sets SB886 and SB890 failed to generate amplification products when initially tested on K8, while alternative primers designed to a homologous insert, SB1366, successfully generated a scorable profile. This suggests that, in future, a number of loci represented in the microsatellite library may be successfully interrogated in the mapping populations if alternative primer sets are designed. However, for some loci,

primer re-design failed to result in amplification, e.g. original primer sets SB321 and alternative primer sets SB606 and SB1300. As discussed in Section 3.4.3, this may have been due to nucleotide diversity existing in the microsatellite flanking regions between the species included in the mapping pedigrees (*S. viminalis* and *S. schwerinii*) and that used for construction of the microsatellite-enriched willow library (*S. burjatica*).

To summarise, development of microsatellite markers required significant investment, both in terms of resources and time. Despite the highly-enriched nature of the microsatellite library, and the large number of unique microsatellite loci detected, development of a sufficient number of polymorphic markers for mapping in K8 was not straightforward.

5.4.2.2 Microsatellite screening

The use of both radioisotope- and fluorescence- based detection strategies were successfully employed for microsatellite genotyping in the K8 population. In some cases, the potential to exploit both of these complementary strategies was beneficial to the study. For example, microsatellite marker SB306 generated a banding pattern that was difficult to score by autoradiography. This was due to the presence of a stutter band, from the larger allele of the two alleles detected, migrating to an equivalent position on the gel as the second true allele. The ability of fluorescence-based systems to more accurately quantify signal strengths and, thus, differentiate between 'true' allele bands and stutter bands, meant that this marker could be scored without difficulty when fluorescent detection methods were employed. In contrast, a number of markers could not be reliably scored when fluorescence was used for detection but were more easily interpreted using autoradiography. This was particularly important for loci at which different microsatellite alleles were preferentially amplified, e.g. SB233, at which the larger of the three alleles detected consistently generated a very faint signal and could not be reliably scored when using fluorescence-based genotyping. The ability to use extended exposure times in autoradiography meant that radioisotope-based methods were more suitable for accurately genotyping loci of this kind.

The use of microsatellite multiplexing strategies can greatly increase sample throughput capacity in microsatellite genotyping studies (Mitchell *et al.*, 1997; Donini *et al.*, 1998). This was clearly the case in this study. For example, alleles at 16 loci could be simultaneously detected for Multiplex E (Table 5.1). This number was comparable to the number of polymorphic AFLP markers that were generated in a single assay by some primer combinations. However, development of viable multiplex combinations was a time-consuming process, and not all of the informative loci available could be incorporated into such strategies due to time limitations. The observation that some of the microsatellite loci were consistently difficult to amplify in multiplex PCR scenarios has been noted by other researchers working with these markers in other willow genotyping studies (S. Trybush, personal communication). Furthermore, a number of loci included in multiplexes generated genotypes that could not be reliably scored for all individuals, resulting in the inclusion of missing data points in the final mapping data set. However, these problems did not offset the benefits of using multiplexing approaches, which were, in general, highly reliable. This study has proved the efficacy of such approaches in genotyping studies in willow. However, further multiplex optimisation may prove beneficial in increasing the efficiency (and for some cases, reliability) with which the available microsatellite markers can be used in future genotyping programmes, both in terms of increasing the number of loci that may be simultaneously screened, and by achieving more consistent signal strengths.

The observation that the DNA sample for K8 individual 318 was in fact derived from *S. viminalis* var. Jorr, highlighted the usefulness of molecular markers in pedigree testing. This error may have arisen from two possible sources: 1) Jorr may have been planted in error at this plot position of the field trial or 2) the leaf material used in DNA extraction may have been mistakenly sampled from a plant contained in the adjacent Jorr plot. Further investigation based on microsatellite analysis of re-sampled material would shed light on the source of this erroneous DNA sample. Time did not allow for this during this present study.

5.4.3 AFLP markers

As observed in the K3 mapping study (Chapter 3), the AFLP technique successfully identified a large number of high-quality, polymorphic markers for mapping in K8. In

general, the technique was highly reproducible, which was of particular importance in this study as PCR assays were performed independently on batches of 80 individuals. However, for a small number of markers that consistently produced very faint signals, several genotypes were not scored in order to exclude potential genotyping errors from the data set. For other cases, where a large proportion of genotypes for a given marker was not scored, this was more a consequence of differences in the quality of resolution between gels, i.e. some gels produced slightly less well-defined banding patterns than others. In such scenarios, the location of some markers very near to an adjacent marker prevented accurate genotyping. Rather than attempt to score such error-prone markers, such questionable genotypes were not included. If sufficient time had been available, then AFLP gels would have been re-run to circumvent this problem and the missing data values filled in.

5.4.4 Bulked segregant analysis

The use of BSA for identification of markers located in a particular map region has been reported in a large number of mapping studies and has also been successfully used in detecting markers linked to a variety of traits in a wide range of species, including a number of trees (Kondo *et al.*, 2000; Moretzsohn *et al.*, 2000; Barros *et al.*, 2002; Cevik and King, 2002). BSA was successfully employed here to identify additional segregating AFLP markers in the vicinity of microsatellite locus SB226. The identification and mapping of markers near to QTL or genes of importance may benefit downstream appplications in a number of ways. First, markers more closely linked to QTL or major genes may be more successfully implemented in marker-assisted selections (Asíns, 2002). Secondly, the availability of several of markers in a region can lead to more accurate positioning of QTL (Kearsey and Farquhar, 1998). Finally, tightly linked markers are a pre-requisite for successful map-based cloning strategies (Tanksley *et al.*, 1995).

The field-based resistant and susceptible DNA pools were included in BSA in an attempt to identify markers putatively linked to the phenotype. As the genetic basis of rust resistance is currently not known in willow, both in terms of the number of genes or loci involved, it was important to identify additional phenotype-linked markers that were not necessarily linked to the SB226 locus. Although several markers were found to segregate between the bulked DNA samples (Table 5.5), the fact that only a single

resistant and a single susceptible bulked DNA sample was used may have resulted in the identification of several false positives. If time had permitted, DNA samples comprising each of the bulks would have been screened individually with the putatively phenotype-linked markers to determine which of the detected markers were true positives. However, time was too restrictive and this line of work was not pursued further.

5.4.5 The K8 parental maps

For both the paternal and maternal linkage maps the number of resolved linkage groups was lower than the haploid chromosome number of willow, with 13 and 12 groups comprising two or more markers being identified, respectively. This indicates that large regions of parental genomes are not represented and additional marker data is required in the future to achieve wider genome coverage. This observation is of consequence for future QTL mapping studies based on the parental maps, as there may be important QTL segregating in regions that are not represented which will, therefore, go undetected. In the paternal map, no regions containing markers displaying distorted segregation ratios were detected, with only a single mapped marker, microsatellite SB337, showing a weakly significant deviance from the expected 1:1 segregation ratio ($P < 0.1$). Similarly, no distorted map regions were identified for the maternal map, although several mapped markers that were unlinked to one another displayed deviance from expected segregation ratios. It is important to note that some slightly distorted segregation ratios may have arisen from biased scoring of difficult markers. For example, for a faint AFLP band, it may be easier to confidently score for the presence of a band rather than its absence. Such scenarios may lead to a bias towards scoring for the presence of the band leading to a false suggestion of a deviant segregation ratio. For this reason, it is important that all possible marker genotypes are scored with equal confidence prior to reporting marker segregation ratios as distorted. The failure to detect any clusters of distorted markers in either parental map, as observed in the published willow maps (Hanley *et al.*, 2002; Tsarouhas *et al.*, 2002), is most probably a consequence of the incomplete map coverage achieved.

5.4.6 The K8 consensus map

As with the parental maps, the number of consensus linkage groups (defined here as those containing at least one marker segregating in both parents) that could be identified (18) was less than the willow haploid chromosome number (19), although the number of linkage groups detected was greater for the consensus map than for the parental maps. This was a consequence of the increased amount of marker segregation information provided by the inclusion of AFLP and microsatellite markers heterozygous in both parents. However, the 18 consensus linkage groups detected may not necessarily reflect 18 chromosomes.

The eight linkage groups comprising markers from a single parent exclusively may be representative of regions linked to those consensus linkage groups detected in this study, but were not observed as linked due to a paucity of markers in potentially connecting regions. Alternatively, these groups may correspond to distinct chromosomes. Furthermore, linkage cannot be determined directly for paternal and maternal markers in the absence of linked markers heterozygous in both parents (Hanley *et al.*, 2002). Hence, it is possible that the exclusive groups of the paternal and maternal maps may, in some cases, be representative of corresponding linkage groups of the consensus map, with linkage remaining undetected due to an absence of linked markers heterozygous in both parents. The generation of additional marker segregation data for inclusion in future linkage analysis should help to resolve all of the above uncertainties.

As in construction of the K3 map (Section 3.3.6), several AFLP markers of segregation type *a0xa0* were problematic and were excluded from linkage analysis for those reasons outlined in Section 3.4.8. Although these markers could have been forced on to the consensus map they were excluded in order to allow construction of a robust consensus map. This was deemed important given that the consensus map would be used as a basis for QTL analysis using interval mapping methods, for which the correct ordering of markers is of prime importance (Asíns, 2002).

AFLP markers showed good distribution across the map, highlighting the usefulness of the technology for rapid map coverage. The clustering of AFLP markers on linkage groups A and B may be a result of suppressed recombination in those regions (Section

3.4.8). Microsatellites also showed good genome distribution, with at least one marker of this type on all but one (Group W♂) of the linkage groups detected. With the exception of linkage group O, which contained two markers displaying distorted segregation ratios, no obvious clusters of distorted markers were detected. As postulated with the parental maps, this may be a consequence of the limited map coverage obtained.

5.4.7 Comparisons with the K3 map

At present, the K3 and K8 consensus linkage maps differ greatly in the number of markers included. Information regarding both the increased density of markers and the greater genome coverage achieved in K3, when compared to K8, may be exploited in future studies to target regions that are not currently represented in the K8 map. While the locus-specific nature of microsatellites makes them amenable to direct transfer from one cross to another, for AFLPs, it is preferable to first convert target markers into locus-specific markers such as Sequence Characterised Amplified Regions (SCARs) (Paran and Michelmore, 1993). Furthermore, as this approach provides single-locus markers for use in subsequent screenings, the high costs associated with the multi-locus AFLP technique may be avoided (Brugmans *et al.*, 2002).

The transferable nature of the microsatellites facilitated the identification of several putatively homeologous linkage groups of the K3 and K8 linkage maps. However, the fact that a significant number of microsatellites that were mapped in K3 could not be mapped on K8, due to absence of polymorphism or complex banding patterns, decreased the certainty of such results. The identification of putatively homeologous groups was based on a single common microsatellite marker in several cases, and should, therefore, be interpreted with caution. This is highlighted by the discrepancy identified regarding the positioning of loci SB274 and SB504 on the same linkage group in K3, but on distinct groups in K8. This observation may have resulted from the duplication of loci within the willow genome. If the polymorphism used for mapping was detected at one locus in the K3 population, but at another corresponding homeologous locus in K8, then the observed result may be explained. Discrepancies in microsatellite positions have also been reported in comparative mapping studies in poplar (Cervera *et al.*, 2001). For example, microsatellite markers PMGC61 and PMGC409 were located on group VIII of *P. nigra* and *P. trichocarpa* maps and also

on the same linkage group in the map of Bradshaw *et al.* (1994). However, on the *P. deltoides* map, these markers were located on distinct linkages groups (VI and VIII, respectively).

Further insight into the relationships between linkage groups of the K3 and K8 linkage maps could be gained through mapping additional markers that segregate in both populations. The microsatellite markers generated in the most recent development phase (SB961 to SB1534) were not tested for their ability to detect polymorphism in K3. Of these, those that were mapped in K8 should be tested in K3 and mapped, if polymorphic, in future studies.

5.5 Chapter summary

- A genetic linkage map based on the large K8 mapping population was constructed for use in QTL analysis.
- Microsatellite markers were successfully employed to identify the two parents of the K8 mapping population. This result was later substantiated during further marker screening for mapping.
- A double pseudo-testcross strategy was employed, resulting in the construction of separate framework parental maps and a preliminary consensus map.
- Molecular marker screening focused on informative, co-dominant microsatellites, however, a limited number of AFLPs were generated to rapidly improve map coverage.
- Additional microsatellite markers were developed, following characterisation of additional microsatellite library inserts.
- To increase genotyping throughput, microsatellite multiplexing approaches based on fluorescent detection were used.
- BSA was used to select AFLP primer combinations that would yield markers in the vicinity of microsatellite marker SB226, a marker that was found to be linked to rust resistance in early non-parametric QTL analysis.
- The preliminary maternal framework map comprised 27 AFLP markers and 24 microsatellites. The resulting 12 linkage groups (including duplets) spanned 528.7 cM with and average interval between markers of 13.6 cM.

- The preliminary paternal framework map comprised 28 AFLP markers and 22 microsatellites. The resulting 13 linkage groups (including duplets) spanned 490.0 cM with an average interval between markers of 13.2 cM.
- The parental linkage maps were integrated to produce a preliminary consensus linkage map comprising 19 *abxaa* AFLP markers, 20 *aaxab* AFLP markers, 49 *a0xa0* AFLP markers and 50 microsatellites. The total map length (including duplets) of these groups was 1191.8 cM with an average interval between markers of 10.1 cM.
- Microsatellite markers allowed the identification of several homeologous linkage groups between the K8 and K3 (Chapter 3) maps. With one exception, microsatellite marker groupings were in agreement with those identified in K3.
- Although the K8 linkage maps will require additional work to improve genome coverage, marker saturation and accuracy in some regions, these maps will be suitable for use in preliminary QTL analyses, as described in Chapter 6.

Chapter 6. Trait analysis and QTL mapping

6.1 INTRODUCTION 206

6.1.1 QTL analysis in outbreeding species 206

6.2 MATERIAL AND METHODS 206

6.2.1 Additional trait assessments 206
- 6.2.1.1 Field-based assessments of rust resistance 207
- 6.2.1.2 Non-destructive yield assessments 207
- 6.2.1.3 Destructive yield assessments 207
- 6.2.1.4 Laboratory-based beetle feeding experiment 207

6.2.2 Examination of spatial variability and adjustment of K8 trait values 208

6.2.3 Phenotypic correlations between characters 211

6.2.4 QTL analysis 211

6.3 RESULTS 214

6.3.1 Additional trait assessments 214
- 6.3.1.1 Field-based trait assessments 214
- 6.3.1.2 Laboratory-based beetle feeding assay 217

6.3.2 Adjustment of K8 trait values for spatial heterogeneity 218

6.3.3 Correlations between characters 219
- 6.3.3.1 Correlations between growth traits in the establishment year 221
- 6.3.3.2 Correlations between growth traits in the second year (2001-2002) 223
- 6.3.3.3 Correlations between growth traits in the third year (2002-2003) 223
- 6.3.3.4 Correlations between traits over three years 224

6.3.4 QTL analysis I. K8 225
- 6.3.4.1 The K8 paternal data set 226
- 6.3.4.2 The K8 maternal data set 229
- 6.3.4.3 The K8 consensus data set 231

6.3.5 QTL analysis II. K3 239

6.3.5.1 The K3 paternal data set 239

6.3.5.2 The K3 maternal data set 242

6.3.5.4 The K3 consensus data set 246

6.4 DISCUSSION 252

6.4.1 Fundamental characteristics of the traits assessed 252

6.4.1.1 Growth traits 252

6.4.1.2 Rust resistance 253

6.4.2 Correlations between characters in K8 254

6.4.3 QTL analysis: Approach 256

6.4.3.1 QTL analysis : K8 258

6.4.3.1.1 General considerations 258

6.4.3.1.2 QTL detection : K8 259

6.4.3.2 QTL analysis : K3 264

6.4.3.2.1 General considerations 264

6.4.3.2.2 QTL detection : K3 264

6.5 CHAPTER SUMMARY 267

6.1 Introduction

The work described in this chapter concerns the use of previously constructed linkage maps (Chapters 3 and 5) and phenotypic data generated in several trait assessments in QTL analysis aimed at the identification of genomic regions underlying variation in traits of agronomic importance. In addition, more fundamental issues regarding relationships between various traits are addressed.

6.1.1 QTL analysis in outbreeding species

As with linkage mapping, QTL analysis in outbred populations is subject to additional complications in comparison to studies based on populations derived from fully homozygous parents due to differences in marker (and QTL) segregation type. As described in Section 2.7.1, this was taken into account in the choice of the JoinMap software for linkage analysis. Similarly, for the QTL mapping reported upon here, the fact that a mapping strategy was a pseudo-testcross approach influenced the methods used for analysis of the data and, in particular, the choice of the QTL analysis software used (see Section 6.2.4).

6.2 Material and Methods

6.2.1 Additional trait assessments

Following the choice of the population for further study, additional trait assessments to those previously described in Chapter 4, were performed. For all field-based trait assessments, the entire K8 progeny (1-947) and the original K3 individuals (1 – 66; those used to construct the K3 linkage map (Chapter 3) were included. Data were recorded separately for plants growing in position 3 and 4 of each plot (Figure 4.1c), unless otherwise stated. To allow for the adjustment of resulting trait data, according to any spatial variation that might be present across the trial site, all reference plots containing *S. viminalis* var. Jorr were also assessed.

6.2.1.1 Field-based assessments of rust resistance

Further assessments of field-based rust resistance were performed in September of both the 2001 and 2002 growing seasons. In both instances, rust scores were assigned to individual plants according to the scoring system described previously in Section 4.2.6.1.

6.2.1.2 Non-destructive yield assessments

Following the second year of growth (2001-2002), stem diameters were recorded during January 2002 for all stems of each assessed plant. These assessments were performed as described in Section 4.2.6. Maximum stem heights were also recorded at this time. After the third year of growth (2002-2003), a further assessment of stem diameters and stem heights was performed. However, unlike in 2002, the heights of all stems per stool were recorded in this year.

6.2.1.3 Destructive yield assessments

In accordance with standard short rotation coppice practise, the field trial was harvested following the third year of growth, in February 2003. The fresh weight of each plant growing in positions 3 and 4 of each plot was first recorded. The remaining four stools of each plot were then harvested and weighed as a single unit. For assessment of percentage dry matter content, all stems in positions 3 and 4 of each plot were separately chipped using a standard garden chipper. Approximately 700 g of chipped material from each was transferred into trays, re-weighed and dried in accordance with the drying protocol outlined in Section 4.2.6.2. Total plant material obtained from each stool could not be included in moisture assessments due to limited oven capacity.

6.2.1.4 Laboratory-based beetle feeding experiment

A second beetle feeding experiment was performed for individuals of the K8 population during August 2001. The experimental protocol was as described in Section 4.2.7.2, with the following exceptions. First, due to a shortage of willow beetles (of any species) at LARS in this year, beetles were collected from a SRC

plantation in Redditch, UK on the day prior to the experiment. Second, the design of the experiment, in terms of the layout and number of leaf discs included in each dish, differed from the first experiment in that: 1) more genotypes (121) were included (c.f. 89 genotypes in the 2000 test) in accordance with a balanced 11 x 11 lattice square design (see Appendix II) and 2) the K477-S3 and K477-R13 parents were included in each dish, at an equivalent position each time. Third, eight beetles were released into each dish (c.f. ten in the 2000 test). Finally, plant material for production of leaf discs was collected from K8 progeny growing in the field trial at LARS. Efforts were made to sample the youngest leaves for which sufficient leaf area was available.

Following the assessment, the percentage leaf area remaining was recorded (c.f. total leaf area remaining in the first experiment) using an Optomax V Image Analyser (Synoptics, Cambridge). As resulting data were recorded on a true percentage scale, with the absolute end point of 100, the closer the means are to 100, the smaller the variability and thus the assumption of approximately equal variation across the whole range is not valid. The appropriate data trans-formation to use with such data is the logit (adjusted to allow for values of 100 being included):

$$log\ it = log_e((\%\text{left} + 0.5)/(100.5 - \%\text{left}))$$

6.2.2 Examination of spatial variability and adjustment of K8 trait values

Data analyses described in this section (6.2.2) were performed and validated by Dr. Stephen Powers, Rothamsted Research, UK. All field-derived K8 trait data (Sections 4.3.3, 4.3.4 and 6.3.1) were analysed to identify and adjust for any spatial trends across the field site that may have contributed to variation in trait values recorded in assessments. The method of Residual Maximum Likelihood (REML) (Patterson and Thompson, 1971; Robinson *et al.*, 1982) was used to fit mixed (involving fixed and random effects) models (Searle *et al.*, 1992) to the trait data, employing the GenStat statistical package (©Sixth Edition, Lawes Agricultural Trust, Rothamsted Experimental Station, 2002). Using theory developed by Gleeson and Cullis (1987), Cullis and Gleeson (1991) and Cullis *et al.* (1998), the most appropriate model to correctly describe the effects of spatial trends for each trait was identified. To this end, the method utilised the trait information provided by the control genotype (var. Jorr), which was planted strategically throughout the design. Changes in model deviance

(Genstat Committee, 1993) were used to assess the significance of extra (spatial) terms in models, these changes being distributed as chi-squared on degrees of freedom equal to the number of extra parameters.

In accordance with the field trial design (Figure 4.1), column (y = 1-52) and row (x = 25-47) co-ordinates were assigned to each plot within the K8 field trial, including the tramline rows (x = 30, 36 and 42) as they complete the spatial arrangement of plants in the design. For all traits except percentage dry matter content, analyses were based on data recorded separately for the two plants in positions 3 and 4 (designated A and B in spatial analysis, respectively) at each (x, y) co-ordinate. This feature was built into each model as a further (fixed) design factor so that all data could be modelled for each variable whilst taking account of the order of sampling plants. For %dry-01 and %dry-03, mean values derived from the two plants assessed were used (see Table 6.1 for explanation of all trait abbreviations). Preliminary analysis of the data revealed that a square root transformation was required for traits Shts-01, Shts-02 and Shts-03. Also, a natural log transformation was required for the FW-03.

The best model for traits MxHt-01, FW-01, DW-01, RR-01, MxHt-02, MxHt-03, MnHt-01, MxDia-01, MxDia-01, MxDia-02, MnDia-01 and MnDia-02 was of the form:

$$d_{ijkl} = \mu + AB_k + Geno_l + (x_{AR(1)} \cdot y_{AR(1)})_{ij} + \varepsilon_{ijkl}$$

where, for data values d_{ijkl}, μ is a constant, AB_k (k = 1, 2 (A or B)) is the effect of the genotype on the right or the left (positions 4 and 3, respectively) and $Geno_l$ (l = 1,..., 952) is the effect of the genotype l (originally denoted 8001,..., 8947) including Jorr, parents and grandparents (l = 948, 949, 950, 951 and 952). For these variables, using just the interaction term (as a random effect), between columns and rows $(x.y)_{ij}$ (i = 25,..., 47 and j = 1,..., 52), provided the best description of the spatial characteristics of the design, together with an autoregressive process of order 1, referred to as AR(1), component for both rows and columns. This means that, moving along columns (x) or rows (y), the description of the variation is of the form:

$$x_i = \varphi_1 x_{i-1} + z_{1i}$$

$$y_j = \varphi_2 y_{j-1} + z_{2j}$$

where φ_1 and φ_2 are the autoregressive parameters and $z_{1i}, z_{2j} \sim N(0, \sigma_z^2)$ are purely random processes, with σ_z^2 as the variance of the processes. Therefore, the spatial variation, or adjusting effect, for row i on the variable modelled, depends on the spatial variation for row $i-1$ and similarly for columns. In other words, in this case, the spatial trend is a gradual change across rows and columns, rows or columns being spatially correlated significant to one step. The final term in the model, ε_{ijkl}, is the error in fitting the observation of genotype l at position k, row x_i and column y_j.

Using the same notation as above, the best model for MnHt-03 and MxDia-03 was of the form:

$$d_{ijkl} = \mu + AB_k + Geno_l + (x_{AR(1)}.y)_{ij} + \varepsilon_{ijkl}$$

as significant autoregressive spatial variation was only apparent up and down columns.

For traits RR-02, FW-03, MnDia-03, Shts-01, Shts-02 and Shts-03, no significant spatial variation was detected, so that the models for these variables were simply of the form:

$$d_{ijkl} = \mu + AB_k + Geno_l + (x.y)_{ij} + \varepsilon_{ijkl}$$

For RR-00, the best description of the spatial variation was given by including the row and column terms separately as random effects with no AR components required. The best model was:

$$d_{ijkl} = \mu + AB_k + Geno_l + x_i + y_j + \varepsilon_{ijkl}$$

Similarly, for %dry-01 and %dry-03 the best model maintained separate independent effects of rows and columns:

$$d_{ijl} = \mu + Geno_l + x_i + y_j + \varepsilon_{ijl}$$

6.2.3 Phenotypic correlations between characters

To examine the linear relationship (positive or negative) between any two traits over the K8 population, phenotypic correlations between all traits assessed in the field were calculated, using all pairs of trait values for all individuals. Calculations were performed using the GenStat statistical package (© Sixth Edition, Lawes Agricultural Trust, Rothamsted Experimental Station, 2002), according to theory described in Mead *et al.* (1993).

To test the significance of each correlation, the *F*-statistic, and its associated probability was determined. Given the probabilistic nature of such significance tests, in which multiple comparison were performed, the probability of type I errors occurring (significant correlations were detected by chance alone), was determined according to:

$$\tilde{p} = 1 - (1 - \alpha)^{\frac{1}{m}}$$

where α was set to 0.05 as a realistically small type I error, and $m = 190$ (the number of correlations). A threshold p-value of 0.0003 resulted.

6.2.4 QTL analysis

QTL analysis was performed for the K8 and K3 populations, as defined in Section 6.2.1. For K8, REML estimates (Section 6.2.2) were used in analyses of field-based traits. For laboratory-assessed traits in K8 and K3, and for K8 nursery heights, unadjusted trait values were used. For K3 field traits, unadjusted trait values were used based on mean trait values derived from plants in positions 3 and 4 of each plot. Furthermore, for K3, phenotypic data generated for rust resistance, mean plot weight, number of shoots per stool and resistance to the leaf roll galling midge *(D. marginemtorquens)* in assessments previously performed in Sweden (I. Åhman, personal communication) were also included in analyses. Details of the protocols used in the Swedish assessments are described in Åhman, (1997). A summary of all traits included in QTL analysis is provided in Table 6.1.

Table 6.1. Summary of traits included in QTL analyses. Details of the timing of assessments and the QTL methods used for analysis (KW = Kruskal-Wallis, IM = interval mapping) for each trait are provided. Unless specified, traits were assessed for both K8 and K3. Codes assigned for each trait are indicated.

Traits	Year	Growth season (in field trial)	Code	Method	Assessment described in section
Nursery height	1999	n/a	NH	KW, IM	n/a
Field-assessed traits					
Rust resistance	2000	1st	RR-00	KW	4.2.6.1
Dry weight	2001	1st	DW-01	KW, IM	4.2.6.2
Fresh weight	2001	1st	FW-01	KW, IM	4.2.6.2
% dry matter content	2001	1st	%dry-01	KW, IM	4.2.6.2
Maximum stem diameter	2001	1st	MxDia-01	KW, IM	4.2.6.3
Mean stem diameter	2001	1st	MnDia-01	KW, IM	4.2.6.3
Maximum stem height	2001	1st	MxHt-01	KW, IM	4.2.6.4
Mean stem height	2001	1st	MnHt-01	KW, IM	4.2.6.4
Number of shoots per stool	2001	1st	Shts-01	KW, IM	4.2.6.3
Rust resistance	2001	2nd	RR-01	KW	6.3.1.1
Maximum stem diameter^{K8}	2002	2nd	MxDia-02	KW, IM	6.3.1.1
Mean stem diameter^{K8}	2002	2nd	MnHt-02	KW, IM	6.3.1.1
Maximum stem height^{K8}	2002	2nd	MxHt-02	KW, IM	6.3.1.1
Number of shoots per stool^{K8}	2002	2nd	Shts-02	KW, IM	6.3.1.1
Rust resistance	2002	3rd	RR-02	KW	6.3.1.1
Maximum stem height	2003	3rd	MxHt-03	KW, IM	6.3.1.1
Mean stem diameter	2003	3rd	MnDia-03	KW, IM	6.3.1.1
Mean stem height	2003	3rd	MnHt-03	KW, IM	6.3.1.1
Maximum stem diameter	2003	3rd	MxDia-03	KW, IM	6.3.1.1
Number of shoots per stool	2003	3rd	Shts-03	KW, IM	6.3.1.1
Fresh weight	2003	3rd	FW-03	KW, IM	6.3.1.1
% dry matter content	2003	3rd	%dry-03	KW, IM	6.3.1.1
Laboratory-assessed traits					
Beetle resistance	2000	n/a	BF-00	KW, IM	4.2.7.2
Beetle resistance^{K8}	2001	n/a	BF-01	KW, IM	6.2.1.4
Rust resistance - LET1	2000	n/a	LET1	KW, IM	4.2.7.1
Rust resistance - LET5	2000	n/a	LET5	KW, IM	4.2.7.1

Table 6.1 continued.

Traits	Year	Growth season (in field trial)	Code	Method	Assessment described in section
Swedish data[K3]					
Fresh weight	1992	n/a	FW-92	KW, IM	Åhman, (1997)
Fresh weight	1993	n/a	FW-93	KW, IM	Åhman, (1997)
Number of shoots per stool	1991	n/a	Shts-91	KW, IM	Åhman, (1997)
Rust Resistance	1993	n/a	RR-93	KW	Åhman, (1997)
Rust Resistance	1995	n/a	RR-95	KW	Åhman, (1997)
Rust Resistance	1996	n/a	RR-96	KW	Åhman, (1997)
Rust Resistance	1997	n/a	RR-97	KW	Åhman, (1997)
Rust Resistance	1998	n/a	RR-98	KW	Åhman, (1997)
Resistance to D. marginemtorquens	?	n/a	DM	KW	Åhman, (1997)

K8: trait assessed in population K8 only
K3: traits assessed in population K3 only
?: not known

All QTL analyses were performed using MapQTL® 4.0 software (Van Ooijen *et al.*, 2002) as this package enables analysis of full-sib families in which up to four QTL alleles may be segregating (MapQTL population type CP). Linkage phases were known from linkage analysis (Chapters 3 and 5). As a first step in QTL identification, the non-parametric rank-sum test of Kruskal-Wallis (KW) (Lehmann, 1975) was used to identify associations between single markers and each trait, with significance threshold of $P < 0.005$ used to declare significance. This was the only approach used with data recorded for field-based rust assessments and resistance to the leaf roll galling midge due to the non-continuous, categorical nature of the scoring systems used (Section 4.2.6.1). For all other traits, interval mapping (IM) (Lander and Botstein, 1989; Van Ooijen, 1992) was then performed based on the two parental maps of both the K8 and K3 populations, followed by analyses based on the consensus maps using an 'all-markers approach' (Knott and Haley, 1992; Maliepaard and Van Ooijen, 1994). Five neighbouring markers were used in such analyses.

For analyses based on both parental and consensus linkage maps, the permutation test (Churchill and Doerge, 1994) within the software package was used to determine LOD significance thresholds at both chromosome-wide and genome-wide levels, based on 1000 permutations and using a significance level of $P < 0.005$ in both cases. QTL with associated LOD scores above the genome-wide significance threshold were considered to be significant. QTL with associated LOD scores that exceeded the chromosome-

wide threshold were considered to be suggestive QTL (Van Ooijen, 1999). All charts were constructed using MapChart 2.1® software (Voorips, 2002). For each QTL identified, ambiguity of map position was indicated by both 1-LOD and 2-LOD support intervals (Van Ooijen, 1992).

6.3 Results

6.3.1 Additional trait assessments

6.3.1.1 Field-based trait assessments

Variation in trait values was recorded for all field-based assessments of the K8 and K3 populations. Results of all assessments are presented as distribution plots in Figure 6.1 (K8) and Figure 6.2 (K3). For K8, these plots are presented to allow comparison with distribution plots following adjustment for environmental factors (Section 6.2.2). Distribution plots based on raw, unadjusted trait data for assessments not included here but used for choosing the QTL population, are provided earlier in the thesis (Figure 4.8). As no adjustment for environment was performed for K3, the distribution plots for K3 are representative of the data used in final QTL analyses.

Figure 4.1. Field-based trait distribution plots : K8.

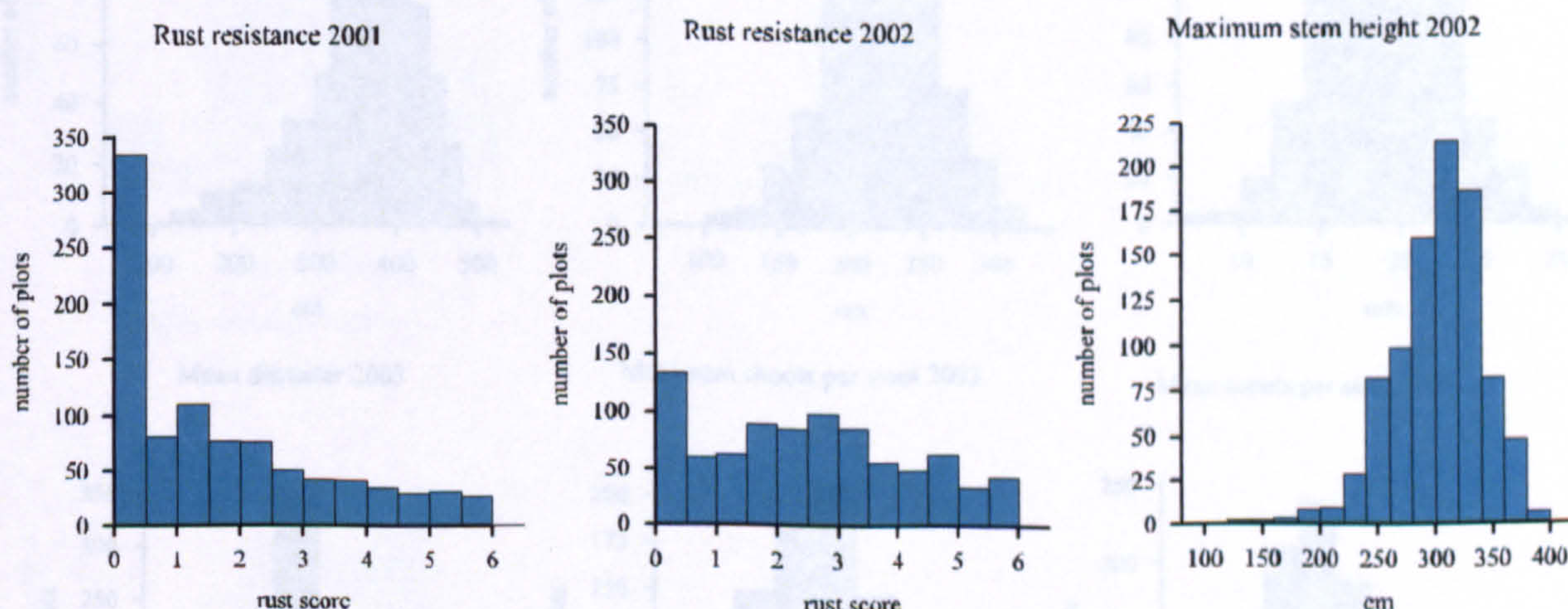


Figure 6.1 continued.

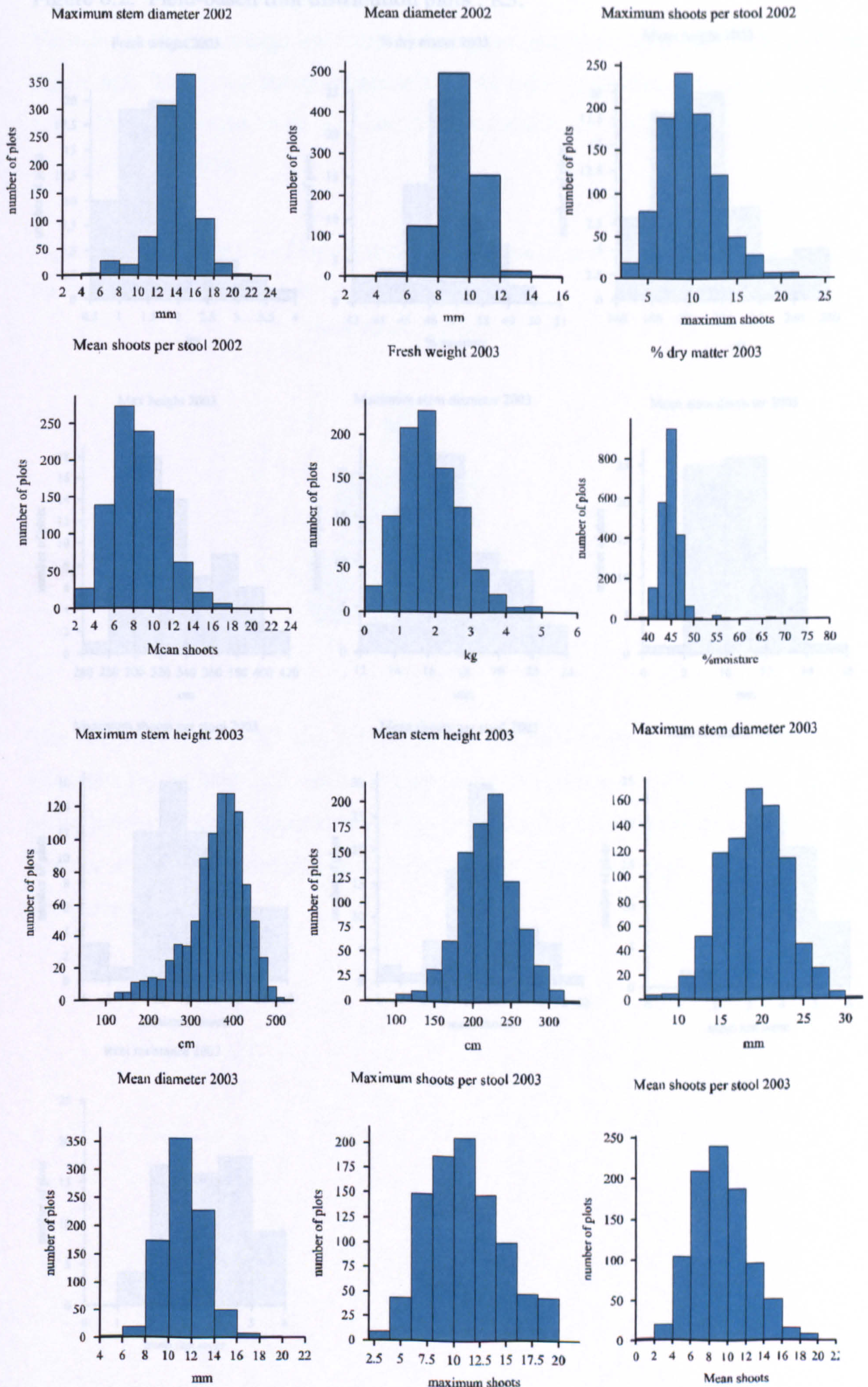
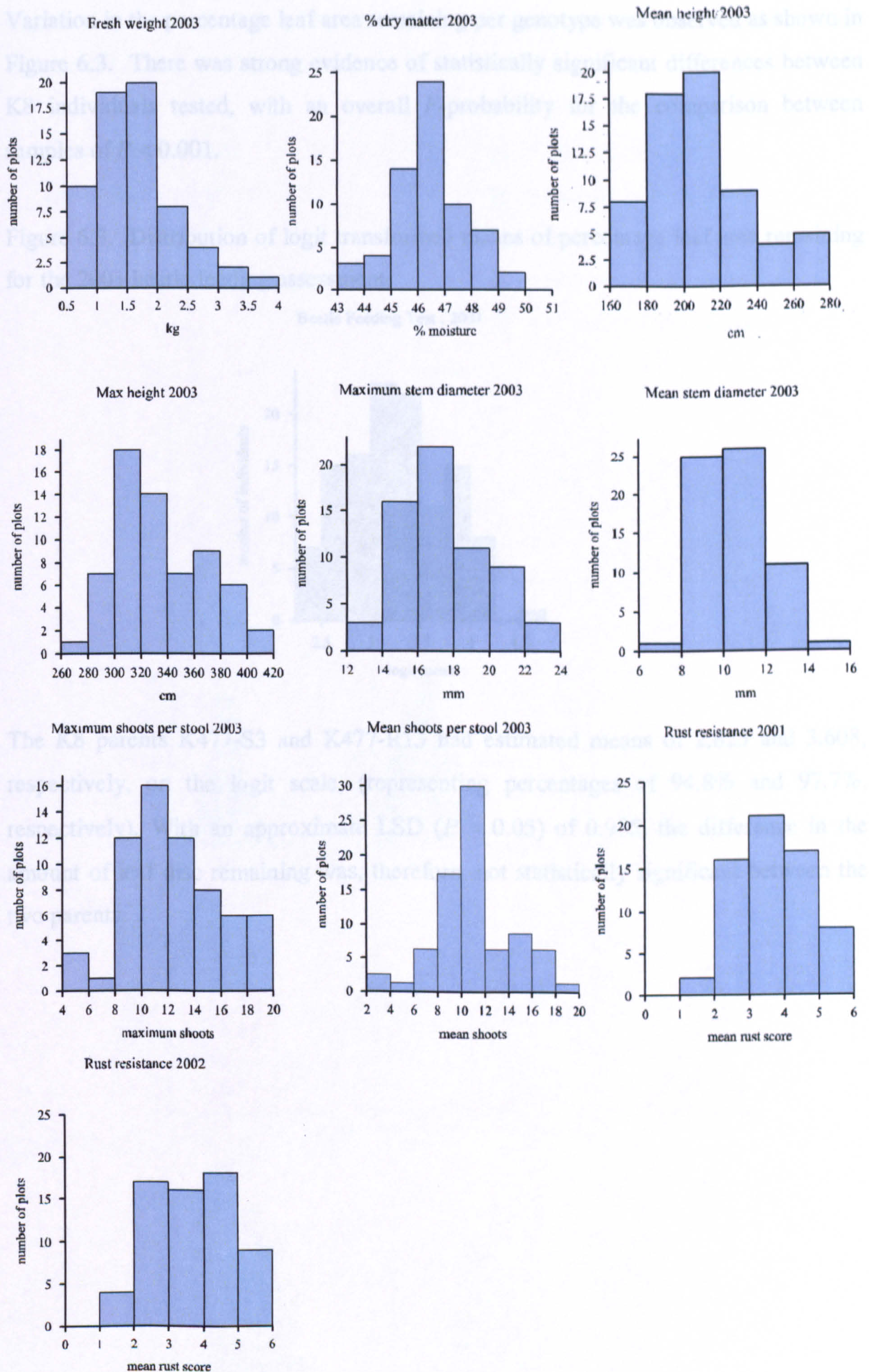


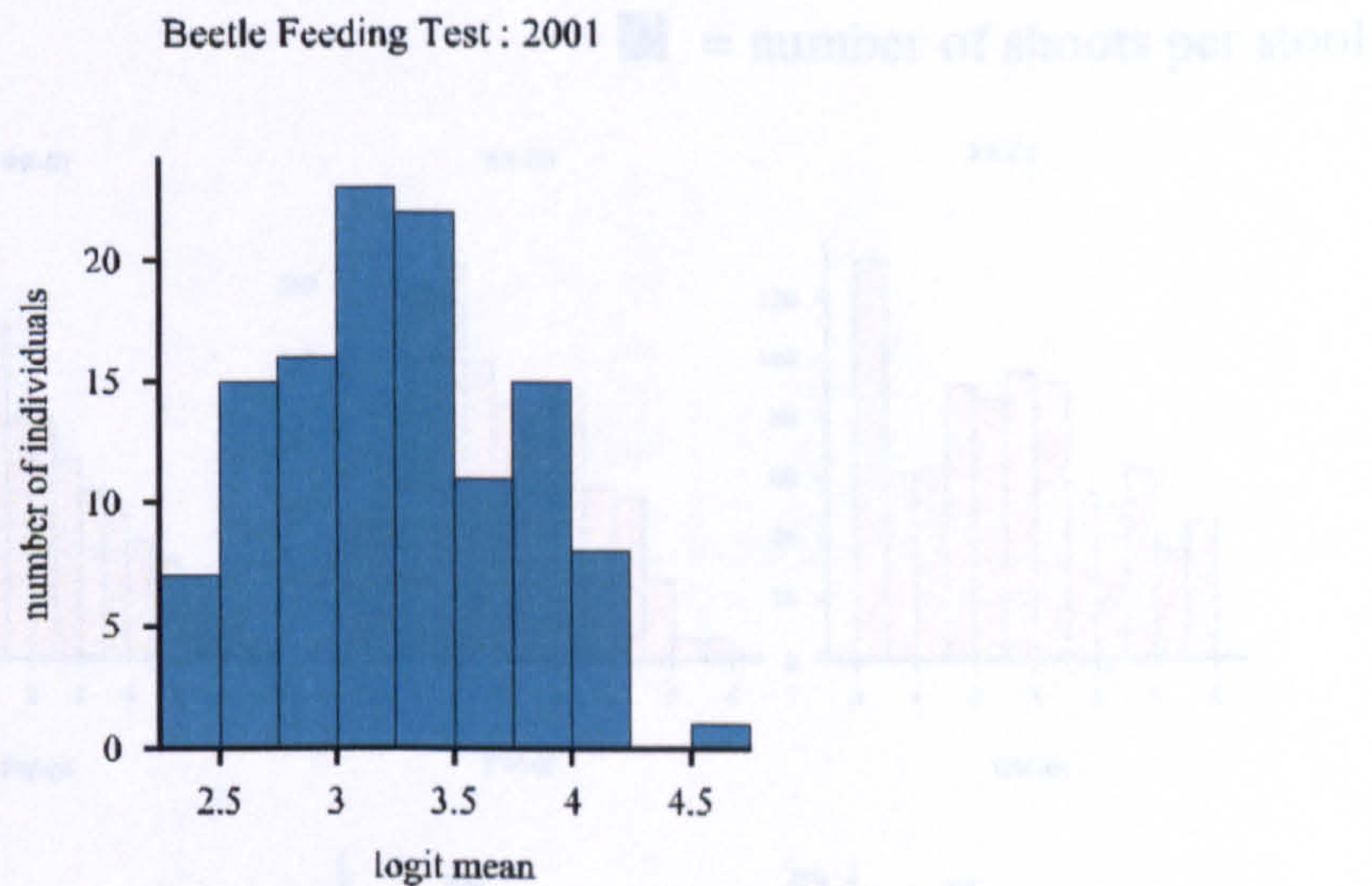
Figure 6.2. Field-based trait distribution plots : K3.



6.3.1.2 Laboratory-based beetle feeding assay

Variation in the percentage leaf area remaining per genotype was observed as shown in Figure 6.3. There was strong evidence of statistically significant differences between K8 individuals tested, with an overall F-probability for the comparison between samples of $P < 0.001$.

Figure 6.3. Distribution of logit transformed means of percentage leaf area remaining for the 2001 beetle feeding assessment.



The K8 parents K477-S3 and K477-R13 had estimated means of 2.825 and 3.608, respectively, on the logit scale, (representing percentages of 94.8% and 97.7%, respectively). With an approximate LSD (P = 0.05) of 0.985, the difference in the amount of leaf disc remaining was, therefore, not statistically significant between the two parents.

6.3.2 Adjustment of K8 trait values for spatial heterogeneity

Figure 6.4. Distribution plots of REML-estimated trait values (x axis) against frequency (y axis) for K8 traits assessed in the field trial.

- = rust resistance
- = fresh weight
- = dry weight
- = maximum stem height
- = mean stem height
- = maximum stem diameter
- = mean stem diameter
- = percentage dry matter
- = number of shoots per stool

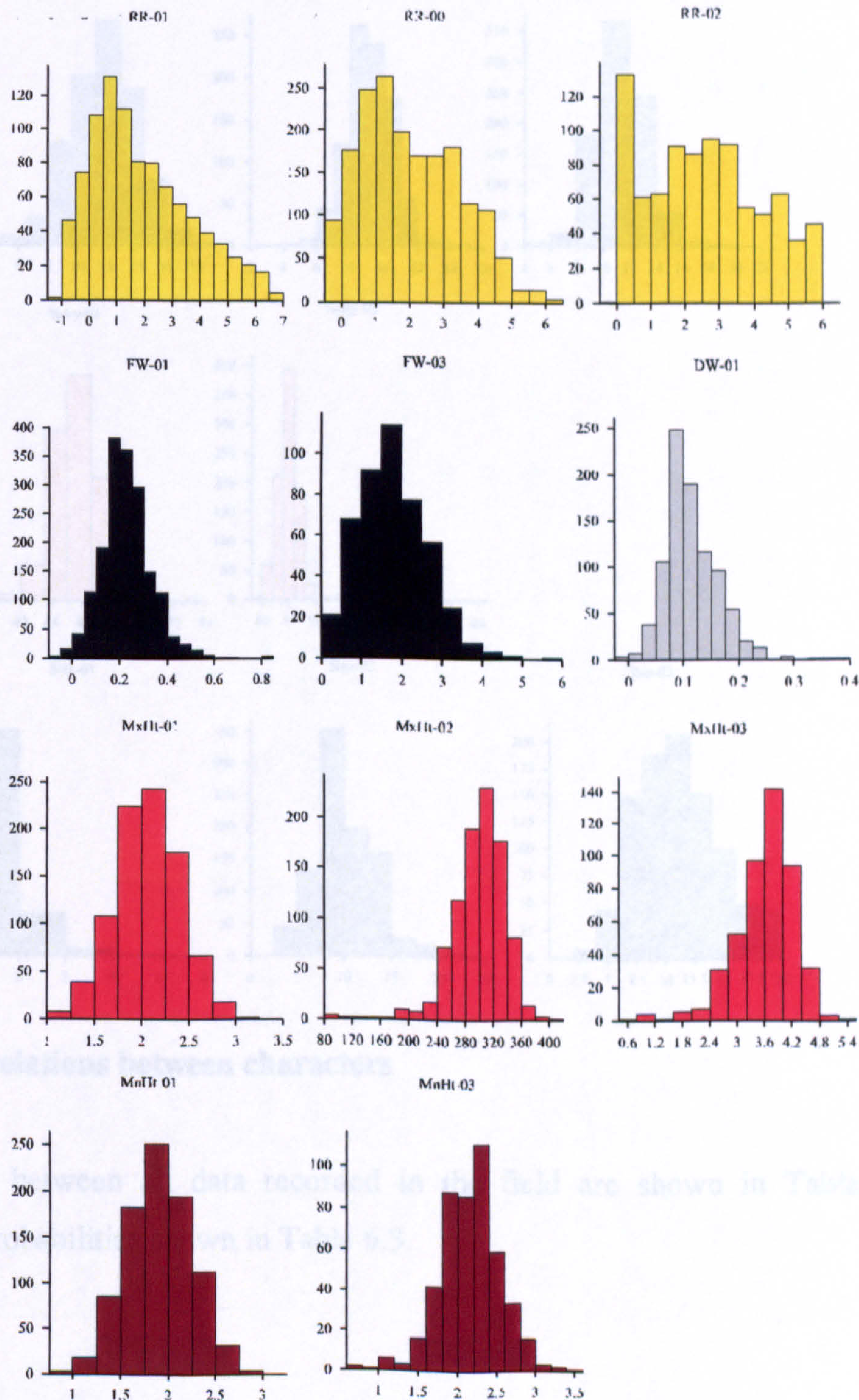
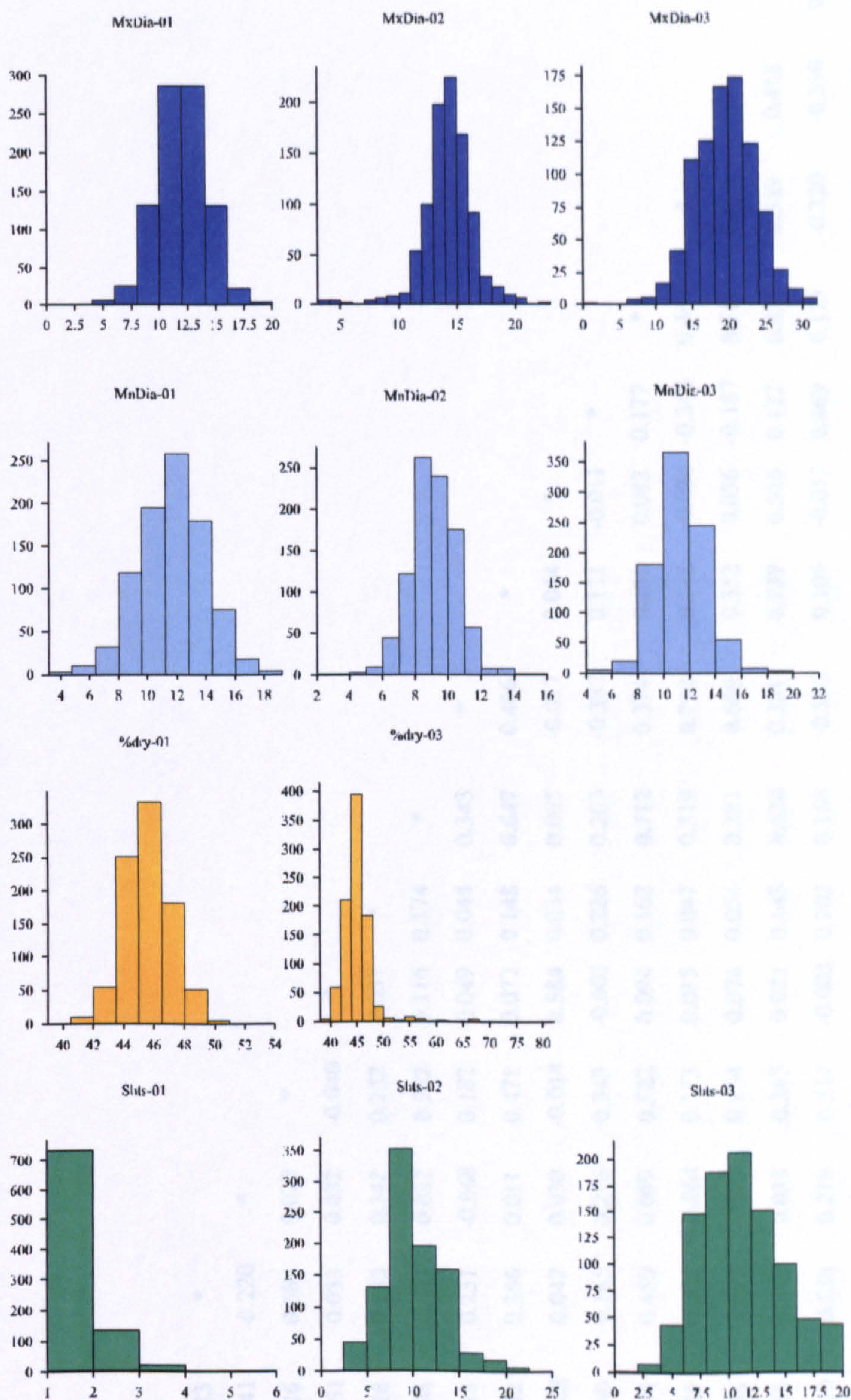


Figure 6.4 continued.



6.3.3 Correlations between characters

Correlations between all data recorded in the field are shown in Table 6.2 with associated probabilities shown in Table 6.3.

Table 6.2. Correlation data for all traits assessed in the K8 field trial. Strong correlations ($r > 0.5$) are shown in *bold* type.

Trait	RR-00	DW-01	MxHt-01	MnDia-01	MnHt-01	Shts-01	MxDia-01	RR-01	FW-01	MxHt-02	MnDia-02	MxDia-02	RR-02	Shts-02	MxHt-03	MnDia-03	MnHt-03	MxDia-03	Shts-03
RR-00	*																		
DW-01	-0.011	*																	
MxHt-01	0.023	**0.742**	*																
MnDia-01	-0.016	0.230	0.287	*															
MnHt-01	0.045	0.187	0.370	**0.583**	*														
Shts-01	0.006	0.327	0.040	-0.041	-0.220	*													
MxDia-01	-0.005	0.254	0.291	**0.946**	**0.584**	0.022	*												
RR-01	0.299	0.032	0.040	-0.051	0.033	0.032	-0.040	*											
FW-01	-0.002	**0.997**	**0.735**	0.228	0.183	0.342	0.252	0.037	*										
MxHt-02	0.050	0.169	0.213	**0.504**	**0.524**	0.032	**0.552**	0.116	0.174	*									
MnDia-02	0.019	0.045	0.065	0.188	0.251	-0.068	0.182	0.049	0.044	0.343	*								
MxDia-02	0.016	0.149	0.154	0.432	0.386	0.011	0.471	0.072	0.148	**0.647**	0.466	*							
RR-02	0.306	0.030	0.052	-0.028	0.042	0.030	-0.014	**0.584**	0.034	0.065	-0.017	0.044	*						
Shts-02	-0.005	0.221	0.116	0.260	0.165	0.216	0.343	-0.003	0.226	0.203	-0.387	0.121	-0.011	*					
MxHt-03	0.024	0.163	0.190	0.474	0.450	0.049	**0.522**	0.094	0.162	**0.712**	0.374	**0.691**	0.082	0.177	*				
MnDia-03	-0.020	0.050	0.065	0.186	0.206	-0.064	0.173	0.015	0.047	0.319	**0.790**	0.444	-0.034	-0.343	0.480	*			
MnHt-03	0.047	0.055	0.109	0.248	0.297	-0.082	0.234	0.074	0.054	0.381	**0.680**	0.352	0.036	-0.187	**0.512**	**0.676**	*		
MxDia-03	-0.027	0.146	0.125	0.352	0.298	0.038	0.385	0.021	0.143	**0.534**	0.381	**0.737**	0.026	0.123	**0.821**	**0.549**	0.418	*	
Shts-03	-0.021	0.197	0.091	0.220	0.138	0.219	0.311	-0.003	0.202	0.188	-0.333	0.106	-0.017	**0.869**	0.173	-0.320	-0.390	0.128	*
FW-03	-0.006	0.254	0.185	0.415	0.356	0.104	0.478	0.030	0.255	**0.549**	0.362	**0.622**	0.006	0.473	**0.746**	0.462	0.445	**0.746**	0.465

Table 6.3. Associated probabilities for correlations. Non-significant correlations ($P > 0.0003$; Section 6.2.3) are shown in *red*.

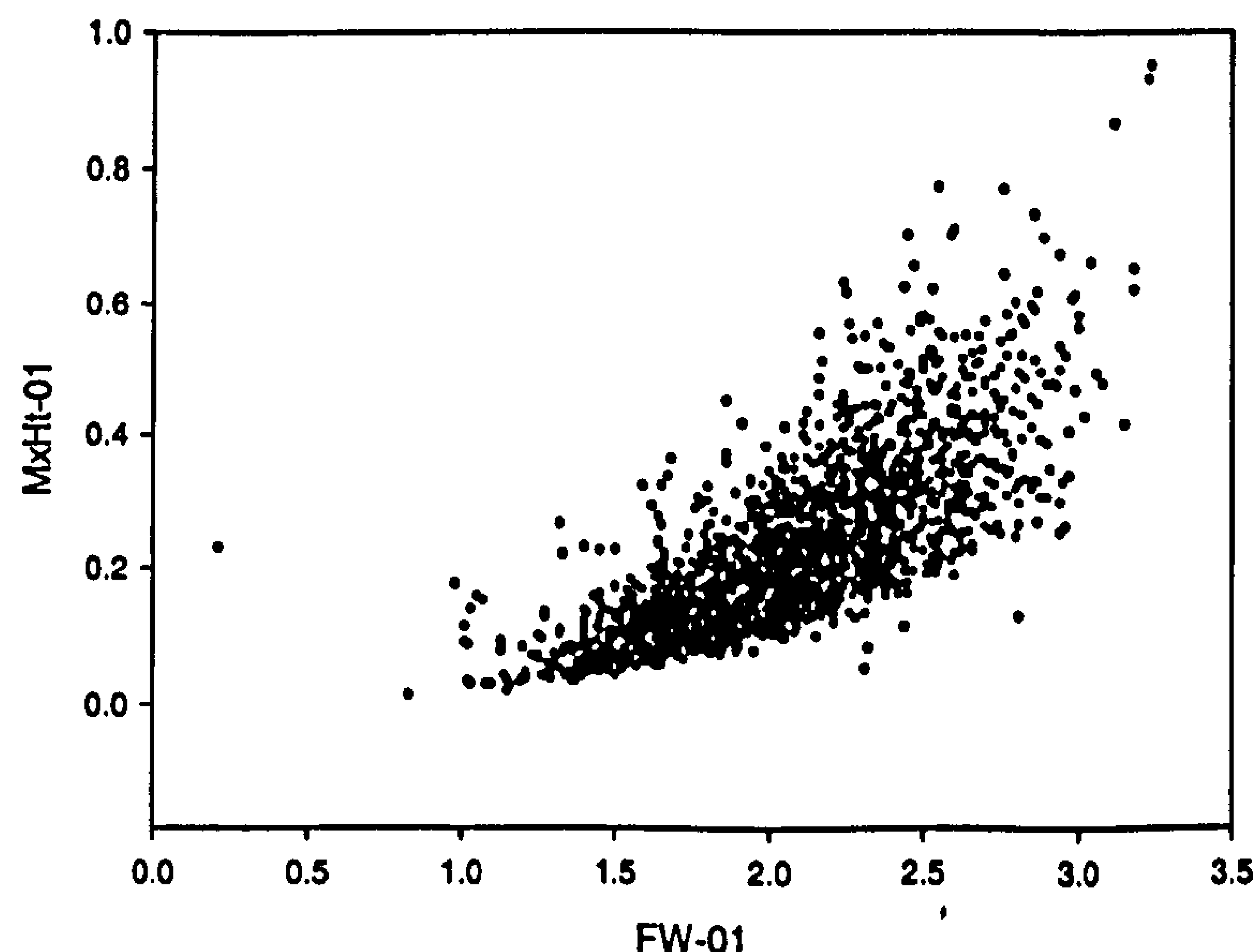
Trait	RR-00	DW-01	MxHt-01	MnDia-01	MnHt-01	Shts-01	MxDia-01	RR-01	FW-01	MxHt-02	MnDia-02	MxDia-02	RR-02	Shts-02	MxHt-03	MnDia-03	MnHt-03	MxDia-03	Shts-03
RR-00	*																		
DW-01	0.651	*																	
MxHt-01	0.348	0.0000	*																
MnDia-01	0.493	0.0000	0.0000	*															
MnHt-01	0.064	0.0000	0.0000	0.0000	*														
Shts-01	0.797	0.0000	0.0953	0.1004	0.0000	*													
MxDia-01	0.847	0.0000	0.0000	0.0000	0.0000	0.3795	*												
RR-01	0.000	0.1884	0.0980	0.0269	0.1606	0.1902	0.0800	*											
FW-01	0.938	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1238	*										
MxHt-02	0.030	0.0000	0.0000	0.0000	0.0000	0.1860	0.0000	0.0000	0.0000	*									
MnDia-02	0.412	0.0609	0.0067	0.0000	0.0000	0.0046	0.0000	0.0302	0.0712	0.0000	*								
MxDia-02	0.495	0.0000	0.0000	0.0000	0.0000	0.6634	0.0000	0.0012	0.0000	0.0000	0.0000	*							
RR-02	0.000	0.2288	0.0371	0.2325	0.0899	0.2289	0.5676	0.0000	0.1770	0.0051	0.4593	0.0581	*						
Shts-02	0.818	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.9038	0.0000	0.0000	0.0000	0.0000	0.6275	*					
MxHt-03	0.303	0.0000	0.0000	0.0000	0.0000	0.0443	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0004	0.0000	*				
MnDia-03	0.404	0.0467	0.0087	0.0000	0.0000	0.0106	0.0000	0.5304	0.0620	0.0000	0.0000	0.0000	0.1564	0.0000	0.0000	*			
MnHt-03	0.041	0.0224	0.0000	0.0000	0.0000	0.0007	0.0000	0.0009	0.0258	0.0000	0.0000	0.0000	0.1180	0.0000	0.0000	0.0000	*		
MxDia-03	0.262	0.0000	0.0000	0.0000	0.0000	0.1264	0.0000	0.3696	0.0000	0.0000	0.0000	0.0000	0.2828	0.0000	0.0000	0.0000	0.0000	*	
Shts-03	0.371	0.0000	0.0002	0.0000	0.0000	0.0000	0.0000	0.8815	0.0000	0.0000	0.0000	0.0000	0.4693	0.0000	0.0000	0.0000	0.0000	0.0000	*
FW-03	0.787	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1825	0.0000	0.0000	0.0000	0.0000	0.7900	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

6.3.3.1 Correlations between growth traits in the establishment year

Several correlations between growth traits were identified upon analysis of the trait data generated in the establishment year. With the exception of any trait with RR-00, and trait Shts-01 with MxHT-01, MnDia-01 or MxDia-01, all correlations were significant ($P < 0.0003$). As may have been expected, a strong correlation ($r = 0.997$) between fresh and dry weight data (FW-01 and DW-01) was observed (Table 6.2).

For potential yield components, strong relationships between both FW-01 and DW-01 and MxHt-01 were observed ($r = 0.735$ and 0.742, respectively). These correlations were higher than those observed when MnHt-01 was compared to FW-01 and DW-01 ($r = 0.183$ and $r = 0.187$). Similarly, with stem diameter data, stronger associations between FW-01 and DW-01 with MxDia-01 were observed ($r = 0.252$ and $r = 0.254$) in comparison to FW-01 and DW-01 with MnDia-01 ($r = 0.228$ and 0.230, respectively). Interestingly, a non-linear relationship between MxHt-01 and yield was identified (Figure 6.5).

Figure 6.5. Scatter plot of MxHt-01 vs. FW-01 showing a non-linear relationship between maximum stem height and fresh weight (2001 data).



Number of shoots per stool (Shts-01) showed no correlation with any other growth traits in this year with the exception of FW-01 and DW-01 for which weak correlations were observed (r = 0.342 and 0.347, respectively)

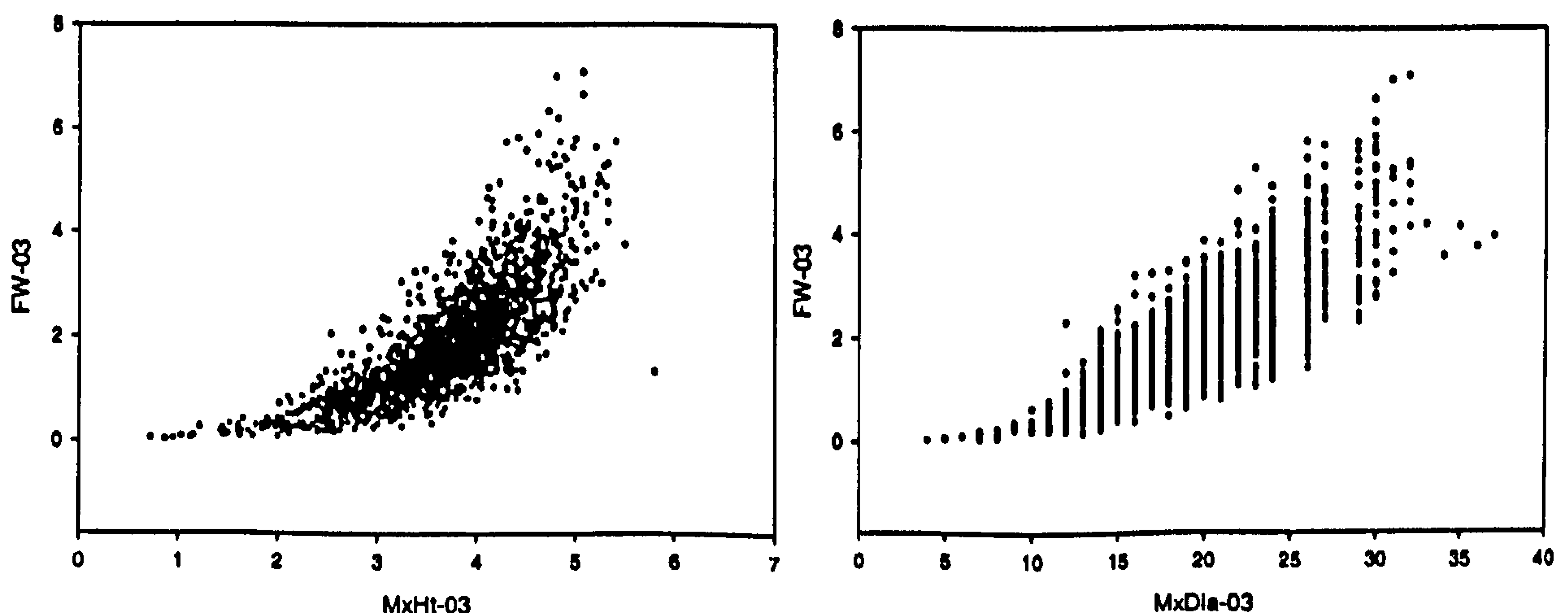
6.3.3.2 Correlations between growth traits in the second year (2001-2002)

In this year (the first year of growth since the previous harvest) all correlations were statistically significant with the exception of RR-01 with MxDia-02 and MnDia-02 and RR-01 with Shts-02. A strong correlation between MxHt-02 and MxDia-02 (r = 0.647) was observed. Weak to moderate correlations between MxHt-02 and MnDia-02 (r = 0.343), MxHt-02 with Shts-02 (r = 0.203) and MxDia-02 with MnDia-02 (r = 0.466), were detected. A weak negative correlation (r = -0.387) between Shts-02 and MnDia-02 was also identified.

6.3.3.3 Correlations between growth traits in the third year (2002-2003)

For this year, most comparisons were significant, although, as in previous years, correlations between number of shoots per stool and all other traits were weaker in comparison to relationships between other traits, and no growth traits were correlated with rust levels. The strongest correlations identified were between MxHt-03 with FW-03 (r = 0.746) and MxDia-03 with FW-03 (r = 0.746). These maximum height and diameter traits were also highly associated with corresponding mean height and diameter traits. Furthermore, height components were highly correlated to diameter components. Similar to data generated in 2001, but in this case more marked, a non-linear relationship between MxHt-03 (and MnHt-03 to a lesser degree) with FW-03 was observed. In contrast to 2001, in 2003 this phenomenon was also evident for MxDia-03 (and MnDia-03) as shown in Figure 6.6.

Figure 6.6. Scatter plots of MxHt-03 vs. FW-03 and MxDia-03 vs. FW-03 showing non-linear relationships between maximum stem height and diameter with fresh weight (2003).



For number of shoots per stool (Shts-03), evidence of a negative relationship with mean diameter (MnDia-03) was detected as in the previous year. In addition, in 2003 a similar negative correlation was identified between this trait and MnHt-03. However, there was some evidence of a correlation (r = 0.465) between the number of shoots (Shts-03) and fresh weight (FW-03).

6.3.3.4 Correlations between traits over three years

Between the establishment year (2000) and the second year (2001 after first harvest) several correlations were observed between traits, particularly for height and diameter traits, e.g. MxHt-02 was strongly related to MnDia-01 (r = 0.504), MnHt-01 (r = 0.524) and MxDia-01 (r = 0.522). For trait data recorded in 2002 and 2003, strong correlations resulted between height and diameter traits, as in the previous comparison. Correlations between both MxHt-02 and MxDia-02 with FW-03 were also convincing (r =0.549 and 0.622, respectively). Between trait data from the establishment year (2001) and that from the third year (before final harvest in 2003), in general, correlations were poorer when compared to comparisons made between other years, with the exception of a convincing association between MxDia-01 and MxHt-03. The strongest correlations between fresh weight (yield) in the final harvest year and establishment year traits were provided by MxDia-01 and MnDia-01 (r = 0.478 and 0.415, respectively). These yield components were more highly correlated with final harvest weights than were the first year (2001) harvest weights.

For number of shoots per stool, the only strong correlation observed was between Shts-02 and Shts-03 (r = 0.869), although there was some evidence of a relationship (r = 0.473) between Shts-02 and FW-03.

Across all three years, correlations amongst rust data were evident, i.e. between RR-00 and RR-01 (r = 0.299), between RR-00 and RR-02 (r = 0.306) and between RR-01 and RR-02 (r = 0.522). For all years, rust was not highly correlated with any other trait.

6.3.4 QTL analysis I. K8

QTL mapping was based on the K8 paternal, maternal and consensus maps as described in Chapter 5. It should be noted that some limitations on the QTL analysis resulted from the fact that there was insufficient time to increase marker coverage on the K8 maps, thus some linkage groups present in one parental map were not represented in the other. In addition, three linkage groups (X, Y and Z) comprised only markers heterozygous in both parents and were not, therefore, included in parental QTL analyses.

Results of QTL mapping based on the K8 paternal, maternal and consensus linkage maps are summarised in Tables 6.4 and 6.5 (Section 6.3.4.1), 6.6 and 6.7 (Section 6.3.4.2), and 6.8 and 6.9 (Section 6.3.4.3), respectively. For each trait, where putative QTL were identified on a given linkage group, the most significant marker-trait associations per group identified by KW analysis are indicated by *asterisks* ($P > 0.1$:*, $P > 0.05$:**, $P > 0.01$:***, $P > 0.005$:****, $P < 0.0001$:*****, $P < 0.0005$:******, $P < 0.00001$:*******). KW results are not shown if the significance was below $P > 0.01$ except where positive results by IM were obtained. Maximum LOD scores for each trait and linkage group are shown if they exceeded the chromosomal significance threshold as determined in permutation tests. Figures *in brackets* indicate the percentage variance explained by each QTL. QTL are indicated in *boxes* when both the KW significance threshold ($P < 0.005$) for KW analysis and the chromosomal LOD significance threshold for IM were exceeded. *Shaded boxes* indicate significant QTL, for which genome-wide significance thresholds for IM analysis were exceeded. Linkage groups were excluded from the tables if there was no evidence of an association with any trait analysed. More comprehensive results of IM analysis are provided in Figures 6.9 and 6.11 for K8 and K3, respectively.

6.3.4.1 The K8 paternal data set

Results of QTL mapping based on the K8 paternal data set are summarised in Table 6.4 for yield-related traits and in Table 6.5 for pest and disease traits. For yield-related traits, 12 significant QTL, located on linkage groups A, I and P, were identified. A further 33 suggestive QTL, significant at the chromosome level in IM, were detected on linkage groups A, B, C, D, G, I, L, P, R, V and W. Of these suggestive QTL, nine were linked to markers that were significant by KW analysis ($P < 0.005$). In several instances of suggestive QTL, co-location with other yield-related traits was observed.

Of the twelve significant QTL mapped, eight were located on linkage group P, for traits FW-01, FW-03, DW-01, MxHT-03, Shts-02, Shts-03, MxDia-01 and MxDia-03. Suggestive QTL for MxHt-01, MxHt-02, MxHt-03, MxDia-02, MnDia-01 and NH were also detected on this group. Similarly, three significant (MxHt-01, Shts-02 and Shts-03) and eight suggestive (FW-01, FW-03, DW-01, MxHt-02, Shts-01, MxDia-01, MnDia-01 and NH) putative QTL were detected on linkage group A. Co-location of several suggestive QTL for related traits (FW-01, DW-01, MxHt-01, MxHt-02 and MnHt-01) was also observed for group C.

For traits assessed over multiple years, consistent putative QTL-linkage group associations were observed in some instances, although the degree of significance of each putative QTL was found to be variable. Linkage group P was associated with fresh weight in both 2001 and 2003 assessments (FW-01 and FW-03), groups A and P were implicated for number of shoots per stool in 2002 and 2003 (Shts-02 and Shts-03) and group P was putatively associated with maximum stem diameter and maximum stem height in all assessment years (MxDia-01, MxDia-02 MxDia-03, MxHt-01, MxHt-02 and MxHt-03).

For field-based rust resistance, KW analysis detected highly significant marker-trait associations on linkage group I for all three years in which assessments were performed. Furthermore, a significant QTL was detected on this group for the rust inoculation test using pathotype LET1. Suggestive QTL for this trait was also identified on linkage groups A and E. For pathotype LET5, IM analysis also identified peaks on groups I and A although, in both cases, the chromosomal significance thresholds were not achieved (data not shown). Similarly, KW analysis of this trait

identified markers putatively linked to this trait although the significance threshold set within this study was not attained for either linkage group. Two putative QTL were identified on groups E and V for field-based resistance in 2000 (RR-00) that were not detected in subsequent years.

A suggestive QTL affecting beetle feeding preference (BF-00) was detected on linkage group V when data generated in the first beetle feeding experiment (2000) (Section 4.3.4.2) were analysed. No QTL were identified following analysis of data produced in the 2001 experiment (BF-01) (Section 6.3.1.2).

Table 6.4. Summarised results of QTL analyses of the K8 paternal (K477-S3) data set for yield-related traits. For explanation see text

LG	FW-01	FW-03	DW-01	MxHt-01	MxHt-02	MxHt-03	MnHt-01	MnHt-03	Shts-01	Shts-02	Shts-03	MxDia-01	MxDia-02	MxDia-03	MnDia-01	MnDia-02	MnDia-03	%dry-01	%dry-03	NH
A	** 1.3 (2)	** 1.23 (2.4)	** 1.71 (2.0)	*** **3.18 (6.2)**					** 1.39 (2.9)	*** **2.39 (5.0)**	**** **3.14 (6.0)**	** 1.93 (3.7)			** 1.6 (2.7)					*** 1.35 (1.5)
B					** 1.68 (1.8)	** 1.17 (1.3)		* 1.1 (1.2)					* 1.45 (1.7)							
C	** 1.58 (2)		** 1.67 (2.1)	** 1.79 (2.6)	**** 1.34 (1.7)		*** 2.18 (3.2)												***	
D						*** 1.62 (1.7)		***						***				*** 2.12 (3.2)		
E																****				
G				** 1.56 (1.7)															***	
I	* 1.75 (2.2)	*** **2.49 (2.9)**	* 1.8 (2.2)											*** 1.48 (1.8)						
L						****							***** 1.42 (1.6)	***						
P	****** **2.97 (3.5)**	******* **3.21 (6.0)**	****** **2.53 (4.4)**	** 1.3 (3.3)	**** 2.01 (3.5)	****** **2.97 (5.0)**		*** 2.16 (2.2)		******* **4.72 (4.6)**	****** **4.62 (4.5)**	***** **2.93 (5.8)**	**** 1.75 (1.8)	******* **2.99 (3.1)**	**** 2.0 (4.0)					**** 1.78 (1.8)
Q																				***
R									** 1.4 (2.1)											
V		** 1.4 (1.5)																		
W									**** 1.49 (1.6)		*** 1.11 (1.1)									

Table 6.5. Summary of putative QTL identified for pest and disease traits in the K8 male parent (K477-R13).

LG	RR-00	RR-01	RR-02	LET1	LET5	BF-00	BF-01
A				***	**		
				1.85 (30.4)			
E	****			*			
	n/a			1.3 (14.9)			
I	*******	*******	*******	****	***		
	n/a	n/a	n/a	**2.4 (26.7)**			
V	****					****	
	n/a					2.52 (22.5)	

6.3.4.2 The K8 maternal data set

Results of QTL mapping based on the K8 maternal data set are summarised in Table 6.6 for yield-related traits and in Table 6.7 for pest and disease traits. For yield traits, six significant QTL were detected on linkage groups A, C, E, K and T, with a further 30 suggestive QTL identified on groups A, B, C, E, F, H, K O, S, T and U. Of these, ten were assigned to linkage groups that contained markers that were also significant for the corresponding trait ($P < 0.005$) in KW analysis.

As with the paternal data, several of the putative QTL were located on linkage group A (FW-03, MxHt-02, Shts-03, %dry-01 and NH). On group K, significant QTL for nursery height and maximum stem height in 2001 were identified. In addition, associations with fresh and dry weight following the establishment year (FW-01, DW-01, MxHt-02, MxHT-03, MnHT-01 and MxDia-02) were also observed for this group. However, for these traits in the second and third years of growth, the significance of these QTL was less than that observed for nursery and establishment years. Similarly, while QTL were detected on linkage group T for FW-01, DW-01 and MxHt-01 and MxHt-02, %dry-01 and NH, little evidence of any QTL effects for this group were observed with the 2003 trait data.

Table 6.6. Summarised results of QTL analyses of the K8 maternal (K477-S3) data set for yield-related traits. For explanation see text.

LG	FW-01	FW-03	DW-01	MxHt-01	MxHt-02	MxHt-03	MnHt-01	MnHt-03	Shts-01	Shts-02	Shts-03	MxDia-01	MxDia-02	MxDia-03	MnDia-01	MnDia-02	MnDia-03	%dry-01	%dry-03	NH
A		**** 1.86 (1.8)			****** 2.06 (2.3)						**** 2.21 (4.4)							**** 2.29 (3.1)	**	********* 2.79 (2.7)**
B																			** 1.41 (1.4)	
C																********* 3.68 (4.9)**	** 1.72 (2.6)	** 2.23 (3.4)		
D																				
E										1.37 (1.9)								****** 3.24 (4.2)**		
F																		* 1.42 (1.9)		
H					*** 1.66 (1.7)												***			
K	**** 1.61 (1.9)		**** 1.68 (2.0)	****** 2.66 (3.2)**	1.06 (1.2)	* 1.31 (1.5)	** 1.35 (1.6)						* 1.11 (1.2)	** 1.34 (1.6)						******** 2.82 (2.8)**
O											** 1.01 (1.5)									
S				**** 1.76 (1.9)			***					*** 1.40 (1.9)			*** 1.55 (2.0)					
T	**** 2.10 (2.3)		**** 2.07 (2.2)	****** 2.46 (2.6)**	**** 1.97 (2.5)				***											*** 1.3 (1.4)
U				** 2.22 (3.1)	*** 1.54 (2.3)		** 1.82 (2.7)					***	***				* 1.22 (1.3)			** 1.38 (1.5)

Table 6.7. Summary of putative QTL identified for pest and disease traits in the K8 female parent (K477-S3).

LG	RR-00	RR-01	RR-02	LET1	LET5	BF-00	BF-01
S				****		****	**
						2.05 (18.2)	1.59 (6.1)
F		***				***	
		n/a				2.22 (21.0)	
K		***	***				
		n/a	n/a				

For field-based rust resistance, two linkage groups (F and K) contained markers that were putatively linked to the phenotype in 2001 and 2002 (RR-01 and RR-02), although the significance threshold of $P < 0.005$ was not achieved in either instance. No evidence for QTL affecting RR-00 was observed. Suggestive QTL were identified for BF-00 on linkage groups F and S, although only the latter reached significance by both KW and IM analyses. For data generated in the second beetle feeding experiment (BF-01), evidence of a suggestive QTL, again on linkage groups S, was observed.

6.3.4.3 The K8 consensus data set

Results of QTL mapping based on the K8 maternal data set are summarised in Table 6.8 for yield-related traits, and in Table 6.9 for pest and disease traits. Several putative QTL were identified when the consensus data set was used that were not detected in analysis of the parental data. Most notably, linkages groups N and Y yielded evidence of several additional QTL. On linkage group N, significant QTL for traits %dry-01 and MnHt-01 were detected, with suggestive QTL identified for DW-01, MxHt-01, MxHt-03, MnHt-03, MnDia-01 and MxDia-03. A significant association with MxDia-01 was also detected on this group by KW analysis, although this was not detected by IM. On linkage group Y, which comprised only markers heterozygous in both parents and was, therefore, not included in parental analyses, significant QTL were detected for MxDia-02, MnDia-02 and MnDia-03. In addition, evidence of suggestive QTL for Shts-01, MxHT-03 and MnHt-03 was observed. Furthermore, some evidence for association (by KW analysis) of markers on this group with FW-03, MxHT-02 and MxDia-03 was detected. Interestingly, with the exception of Shts-01, no QTL were detected on this group for any yield-related traits in the establishment year (2001).

Significant QTL were detected on linkage group B for percentage dry matter content in both years in which destructive assessments were performed (%dry-01 and %dry-03). These QTL were not detected in parental analyses. Additional significant QTL for %dry-01 and %dry-03 were detected on linkage groups N and K, respectively. Once more, these associations were not detected in either of the parental analyses.

Table 6.8. Summarised results of QTL analyses of the K8 consensus data set for yield-related traits. For explanation see text.

LG	FW-01	FW-03	DW-01	MxHt-01	MxHt-02	MxHt-03	MnHt-01	MnHt-03	Shts-01	Shts-02	Shts-03	MxDia-01	MxDia-02	MxDia-03	MnDia-01	MnDia-02	MnDia-03	%dry-01	%dry-03	NH
A	***	****	**	**** 3.54 (6.4)	******	**				**** **4.24 (7.6)**	**** **4.89 (8.3)**					****		******* **4.04 (5.0)**		******* 3.44 (3.8)
B																		******* **6.04 (7.4)**	******* **5.5 (5.9)**	** 3.09 (6.3)
C		*** 2.92 (5.7)			*** 2.6 (4.9)	**** 4.16 (7.6)			***				****	*** 3.2 (5.3)	***	****** **4.23 (5.0)**		*** 2.75 (3.6)		* 2.73 (5.3)
D						***		***				***		*** 2.66 (4.9)				***		
E						****				* 3.27 (5.6)	* 3.56 (6.8)							**** 3.96 (5.2)		
F																		** 3.46 (7.7)		
G								***								***			*** 3.62 (5.7)	
H					*** 3.67 (21.7)			***									***			
I		*** 3.09 (3.3												***						
J		** 3.09 (8.9)								** 2.4 (4.9)									****	
K	****		****	**** 2.7 (5.3)	2.53 (9.7)	* 3.0 (13.2)	** 2.87 (47.1)												******* **5.08 (5.4)**	****** **5.37 (36.0)**
L					***	****							*****	***						
M																				***** **4.68 (7.7)**

Table 6.8 continued.

LG	FW-01	FW-03	DW-01	MxHt-01	MxHt-02	MxHt-03	MnHt-01	MnHt-03	Shts-01	Shts-02	Shts-03	MxDia-01	MxDia-02	MxDia-03	MnDia-01	MnDia-02	MnDia-03	%dry-01	%dry-03	NH
N			** 2.02 (4.1)	***** 2.44 (3.2)		**** 2.52 (3)	******* **4.65 (7.0)**	**** 2.44 (2.4)				******		** 2.35 (13.1)	******* 3.77 (5.3)			***** **3.71 (4.2)**		
O		* 2.69 (19.4)									** 2.30 (16.9)									
Pp	****** 2.97 (3.5)	******* 3.32 (3.4)	****** 2.70 (3.0)	** 1.71 (3.3)	**** 2.01 (3.5)	****** 2.97 (4.9)		*** 2.16 (2.2)		******* **4.72 (4.6)**	****** **4.61 (4.5)**	***** 2.93 (5.8)	**** 1.75 (1.8)	******* 2.99 (3.1)	**** 2.00 (4.0)					**** 1.78 (1.8)
Qp																				***
Sm				**** 1.76 (1.9)			***					***			*** 1.54 (2.0)					
Tm	**** 2.10 (2.3)		**** 2.07 (2.2)	**** 2.46 (2.6)	**** 1.97 (2.5)				***											***
Um				** 2.22 (3.2)	*** 1.54 (2.3)		** 1.82 (2.7)						***							** 1.38 (1.6)
Vp		** 1.4 (1.5)																		
Wp									**** 1.49 (1.6)		*** 1.11 (1.1)									
X																				
Y		***			***	****** 3.3 (3.7)		***** 3.01 (3.5)	*** 2.27 (3.2)				******* **4.69 (5.3)**	****		******* **6.33 (7.0)**	******* **5.15 (5.9)**			
Z								***										****** 2.36 (2.5)	** 1.74 (1.9)	**** 2.69 (3.3)

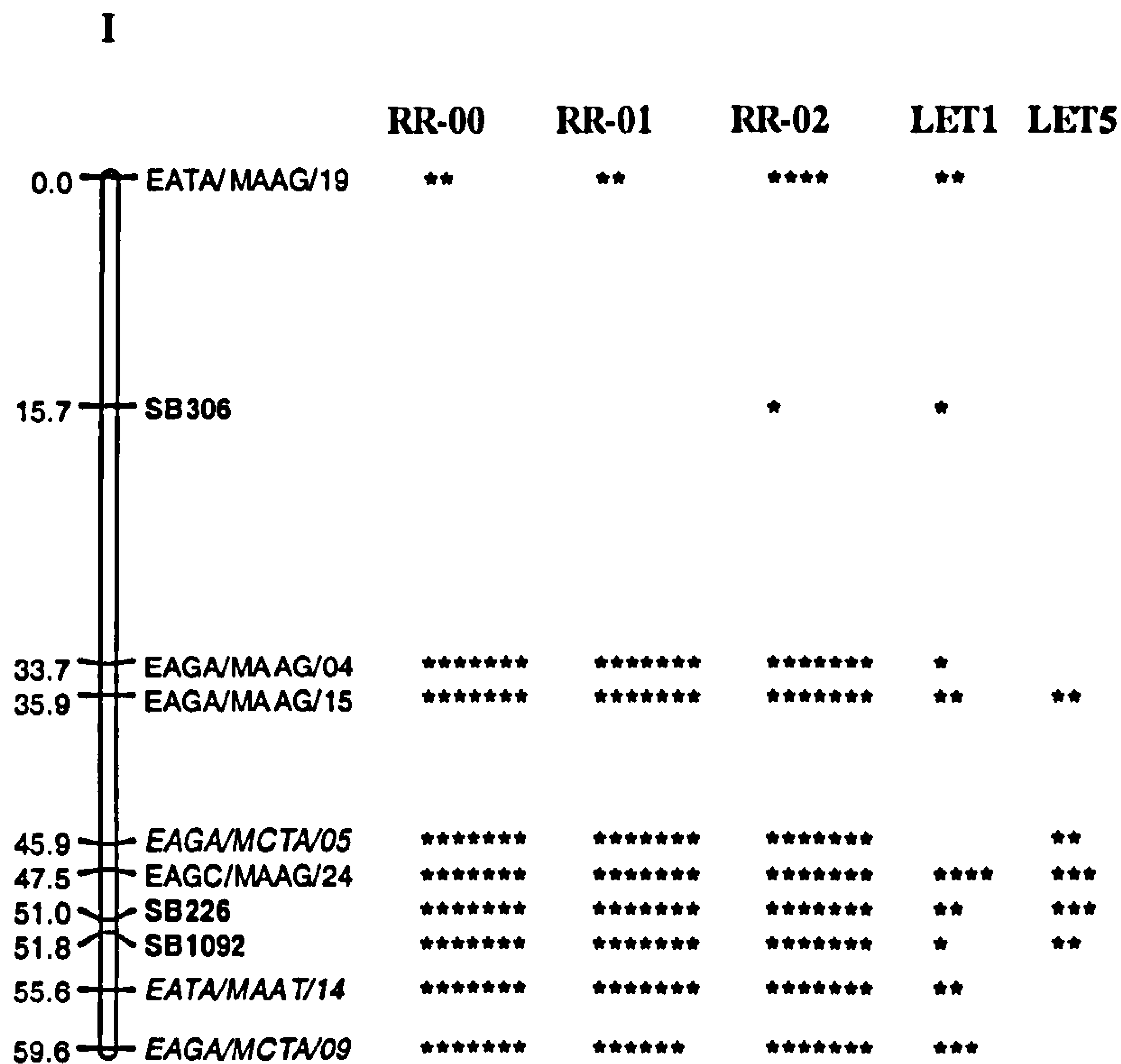
Table 6.9. Summarised results of QTL analyses of the K8 consensus data set for disease and pest traits. For explanation see text.

LG	RR-00	RR-01	RR-02	LET1	LET5	BF-00	BF-01
A	***			***			
	n/a						
B	*******	****	****	***			
	n/a	n/a	n/a				
E	****						
	n/a						
F		***	***			***	
		n/a	n/a				
I	*******	*******	*******	****	***		
	n/a	n/a	n/a				
K		***	***				
		n/a	n/a				
L			****				
			n/a				
M			***				
							2.3 (20.3%)
Pp							
Sm				****		****	**
						2.05 (18.2)	1.58 (6.1)
Vp	****						
	n/a						
X	***						
	n/a						
Z			****				
			n/a				

Evidence for the presence of QTL for rust resistance, from both field and laboratory data, was observed for several linkage groups, although the significance threshold for KW analysis ($P < 0.005$) used in this study was not achieved in all instances As with the paternal data, strong associations ($P < 0.0001$) were detected by KW analysis for markers on linkage group I (not represented in the maternal map) for field-based rust resistance in all three years assessed. Furthermore, this linkage group also showed some evidence of involvement in resistance to both pathotypes LET1 and LET5 in inoculation experiments, although for LET5, criteria for declaring significant QTL

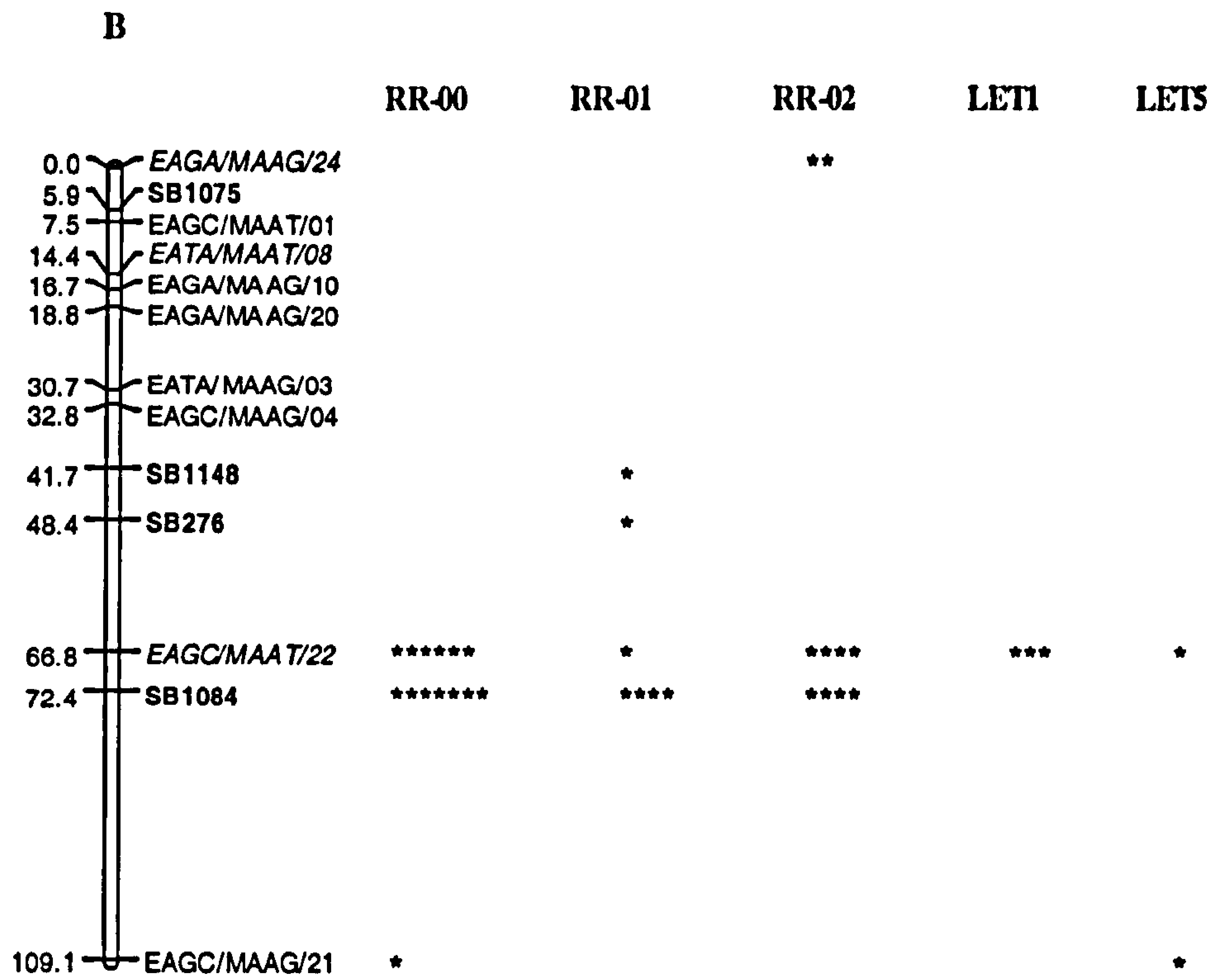
were not met for either KW or IM analyses. Similarly for LET1, while KW analysis detected a significant association, IM analysis failed to do so. More detailed results of KW analysis are shown in Figure 6.7.

Figure 6.7. Results of Kruskal-Wallis analysis for rust traits on consensus linkage group I. The significance of each marker-trait association is indicated by *asterisks* (*: $P < 0.1$, **: $P < 0.05$, ***: $P < 0.01$, ****: $P < 0.005$, *****: $P < 0.001$, ******: $P < 0.0005$, *******, $P < 0.0001$). Marker nomenclature is as described in Figure 3.6.



A second linkage group, group B, also showed evidence of containing a QTL involved in rust resistance, as KW analysis results were significant for the field data in all assessment years. Sub-significant associations were also detected on this group for resistance to LET1 and LET5 in the laboratory-based experiments (Figure 6.8).

Figure 6.8. Results of Kruskal-Wallis analysis for rust traits on K8 consensus linkage group B. The significance of each marker-trait association is indicated by *asterisks* (*: $P < 0.1$, **: $P < 0.05$, ***: $P < 0.01$, ****: $P < 0.005$, *****: $P < 0.001$, ******: $P < 0.0005$, *******, $P < 0.0001$). Marker nomenclature is as described in Figure 3.6.



For beetle resistance, apart from the putative QTL previously identified on maternal linkage group S, no QTL that were significant by KW or IM analyses were detected.

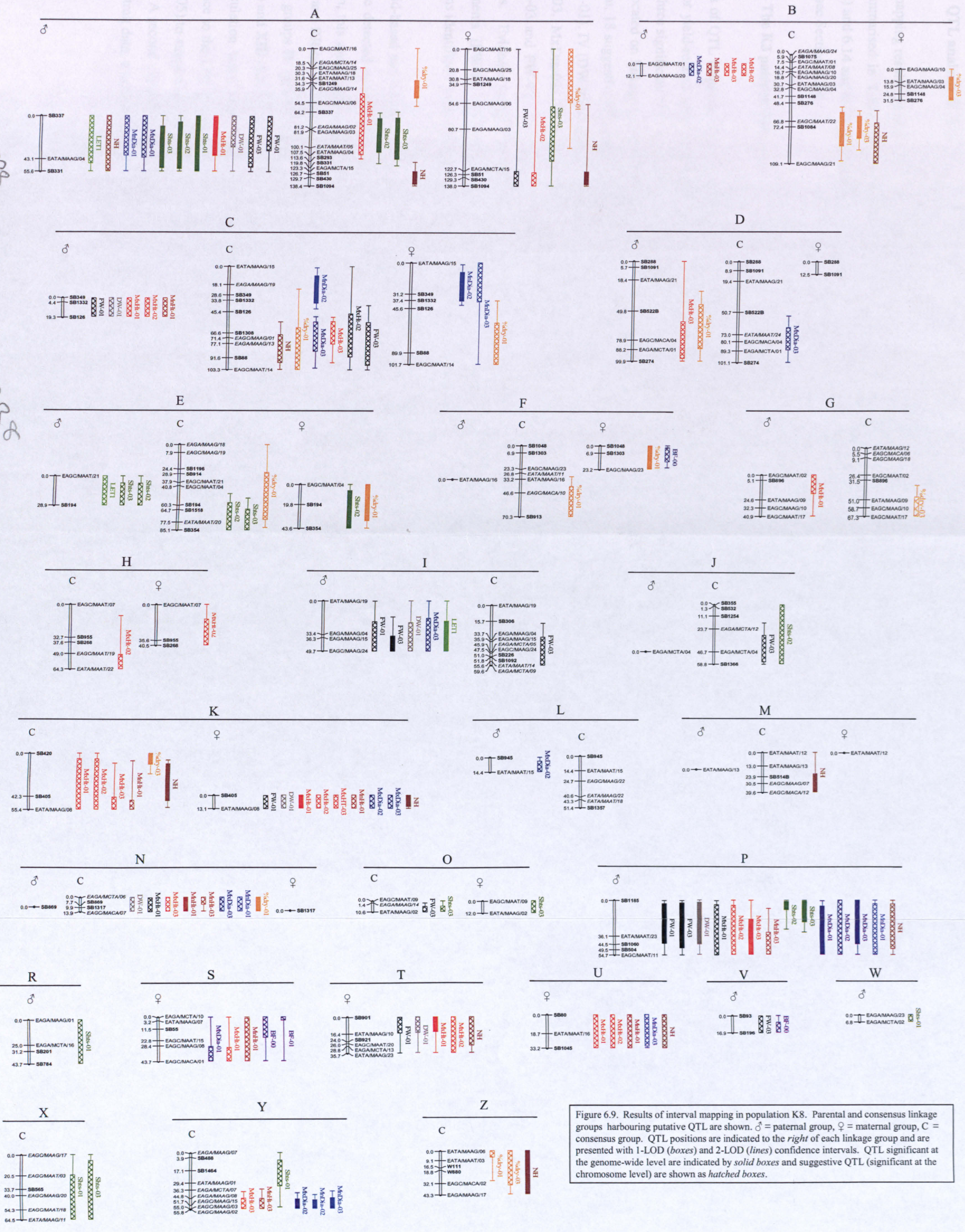


Figure 6.9. Results of interval mapping in population K8. Parental and consensus linkage groups harbouring putative QTL are shown. ♂ = paternal group, ♀ = maternal group, C = consensus group. QTL positions are indicated to the *right* of each linkage group and are presented with 1-LOD (*boxes*) and 2-LOD (*lines*) confidence intervals. QTL significant at the genome-wide level are indicated by *solid boxes* and suggestive QTL (significant at the chromosome level) are shown as *hatched boxes*.

6.3.5 QTL analysis II. K3

QTL mapping results, based on the K3 paternal, maternal and consensus linkage maps, are summarised in Tables 6.10 and 6.11 (Section 6.3.5.1), 6.12 and 6.13 (Section 6.3.5.2) and 6.14 and 6.15 (Section 6.3.5.3), respectively. For an explanation of these tables see Section 6.3.4.

6.3.5.1 The K3 paternal data set

Results of QTL mapping based on the K3 paternal data set are summarised in Table 6.10 for yield-related traits and in Table 6.11 for pest and disease traits. For yield traits, three significant QTL were detected, of which two (for MnHt-03 and MnDia-03) were located on linkage group XIII, and the third (for Shts-91) was on group XV. In addition, 18 suggestive QTL were detected on groups I (%dry-03), III (MxDia-01 and MnDia-01), IV (DW-01 and %dry-01), VIII (FW-92, FW-93, Shts-91), IX (MxHT-03, MnHt-03, MxDia-03), X (Shts-91), XVI (MnDia-03), XVIII (MnHt-03, Shts-03, MnDia-03 and FW-92). Of these, seven were also significant ($P < 0.005$) in KW analysis. Evidence for QTL was obtained upon analysis of data generated in Swedish assessments (FW-92, FW-93 and Shts-91) however, these did not co-locate to positions identified for corresponding traits with the LARS data.

For field-based rust resistance, consistent associations with markers on linkage group III were detected upon analysis of data generated in both Sweden and the UK. In addition, this result was observed over several years ($P < 0.005$ for RR-95, RR-96, RR-97 and RR-00; $P < 0.01$ for RR-01). Additional putative QTL were identified on linkage groups IV (RR-97, RR-01 and RR-02), V (RR-97 and RR-01), VI (RR-96 and RR-01) and XIII (RR-95, RR-96, RR-01 and RR-02). Upon analysis of data produced in inoculation tests, linkage groups III, IV and VI all showed association with resistance to the LET5 inoculum. Furthermore, KW analysis detected slight evidence ($P < 0.05$) to support the involvement of a locus (or loci) on group III in resistance to LET1. A second putative QTL for LET1, which was not detected with any other rust-related trait data, was identified on group Am.

A suggestive QTL for BF-00 was detected on linkage group III. Evidence for the presence of a QTL affecting susceptibility to *D. marginemtorquens* was observed on linkage group XXI. The three markers comprising this group all displayed strong linkage to this trait ($P < 0.0001$ for all markers).

Table 6.10. Summarised results of QTL analyses based on growth traits for the K3 male parent (SW980082 var. Orm). See text for explanation

LG	FW-01	FW-03	DW-01	%dry-01	%dry-03	MxHt-01	MxHt-03	MnHt-03	Shts-01	Shts-03	MxDia-01	MxDia-03	MnDia-01	MnDia-03	FW-92	FW-93	Shts-91
I					*** 2.06 (16.4)												
IIa				***	***												
III											***** 2.48 (16.3)		**** 2.19 (14.6)				
IV			1.68 (11.4)	***** 2.34 (15.6)													
V															***		
VI						***											
VIII															*** 2.21 (15.2)	***** 2.65 (17.4)	** 1.81 (12.3)
IX							**** 1.7 (12.0)	** 1.4 (10.0)			***	*** 1.7 (12.2)					
X																	** 1.74(17.1)
XIII								******* **3.17 (21.6)**						****** **3.24 (21.9)**			
XV																	******* **3.66 (23.2)**
XVI														** 1.11 (8.2)			
XVIII								**** 2.46 (16.4)		**** 1.78 (12.0)				** 1.45 (10.0)	* 1.45 (10.2)		
XX											** 1.49 (10.1)						

Table 6.11. Summarised results of QTL analyses based on pest and disease traits in the K3 male parent (SW980082 var. Orm). See text for explanation.

LG	RR-93	RR-95	RR-96	RR-97	RR-98	RR-00	RR-01	RR-02	DM	BF00	LET1	LET5
IIB		***										
III		*****	*****	******		****	**			****	**	***
		n/a	n/a	n/a		n/a	n/a			1.79 (12.0)		1.81 (12.8)
IV				****			***	***				***
				n/a			n/a	n/a				
V				****			****					
				n/a			n/a					
VI			****				****					***
			n/a									2.21 (17.5)
IX				***								
				n/a								
X		***										
		n/a										
XIII		****	***				***	***				
		n/a	n/a				n/a	n/a				
XXI									*******			
									n/a			
Am											****	
											2.01 (13.5)	

6.3.5.2 The K3 maternal data set

Results of QTL mapping based on the K3 maternal data set are summarised in Table 6.12 for yield-related traits and in Table 6.13 for pest and disease traits. Two significant QTL were detected on linkage groups IB (FW-92) and VIII (Shts-03). Additional suggestive QTL were detected on groups IB (FW-93 and Shts-91), IIA (FW-01, DW-01 and MxHT-01, MxHT-03), IIC (MnHt-03 and MnDia-03), III (%dry-03 and FW-93), IV (MxDia-01 and FW-92), V (%dry-03), VII (Shts-03), XII (%dry-01), XIII (%dry-03 and MxHt-01), XVIII (Shts-03), Af (MnHt-03 and Shts-91), B, (DW-01 and FW-93) and F1 (FW-01 and Shts-03).

For field-based rust resistance, evidence of a QTL on linkage group III was detected by KW analysis, although in 1995 and 2002 the KW significance threshold used in this study ($P < 0.005$;****) was not exceeded ($P < 0.01$ and $P < 0.05$ in 1995 and 2002, respectively). This group was also highlighted as containing a QTL for resistance to LET5 in the inoculation experiment. However, no evidence for an association with the

LET1 pathotype was evident. A second fairly consistent QTL for rust resistance was identified on linkage groups V, although once more, the significance of this association varied over the years. Although not significant by either KW analysis ($P < 0.01$) or IM analysis, there was slight evidence of an association between this linkage group and resistance to the LET5 pathotype. As for group III, no significant association with pathotype LET1 was detected for this group V. Suggestive QTL for resistance to LET1 were, however, observed on groups XVIII and F2.

No significant associations were detected on analysis of data produced in the beetle feeding experiment or the assessment of susceptibility to *D. marginemtorquens*.

Table 6.12. Summarised results of QTL analyses based on growth-related traits in the K3 female parent (SW980084). For explanation see text.

LG	FW-01	FW-03	DW-01	%dry-01	%dry-03	MxHt-01	MxHt-03	MnHt-03	Shts-01	Shts-03	MxDia-01	MxDia-03	MnDia-01	MnDia-03	FW-92	FW-93	Shts-91
IA					***												
IB															****** **4.79 (36.6)**	** 1.53 (12.2)	**** 2.81 (21.1)
IIA	**** 2.15 (14.3)		**** 2.10 (14.0)			**** 2.19 (14.6)	1.31 (9.0)										
IIC								** 1.39 (9.7)						** 1.45 (10.0)			
III					**** 2.03 (13.6)											**** 1.6 (10.9)	
IV											*** 1.55 (11.5)				**** 1.69 (11.8)		
V					**** 2.37(15.7)										***		
VII										*** 1.26 (9.2)							
VIII										****** **3.09 (25.2)**							
XII				**** 1.58 (11.0)													
XIII					** 1.65 (11.2)	***** 2.88 (18.7)											
XV					**** 2.32 (15.4)												
XVIII										**** 2.84 (23.0)							

Table 6.12 continued.

LG	FW-01	FW-03	DW-01	%dry-01	%dry-03	MxHt-01	MxHt-03	MnHt-03	Shts-01	Shts-03	MxDia-01	MxDia-03	MnDia-01	MnDia-03	FW-92	FW-93	Shts-91
AF								***									**
								1.97 (15.8)									1.73 (11.7)
B																***	
			1.16 (8)													1.45 (9.9)	
F1										**							
										1.55 (14.7)							

Table 6.13. Summarised results of QTL analyses based on pest and disease traits in the K3 female parent (SW980084). For explanation see text.

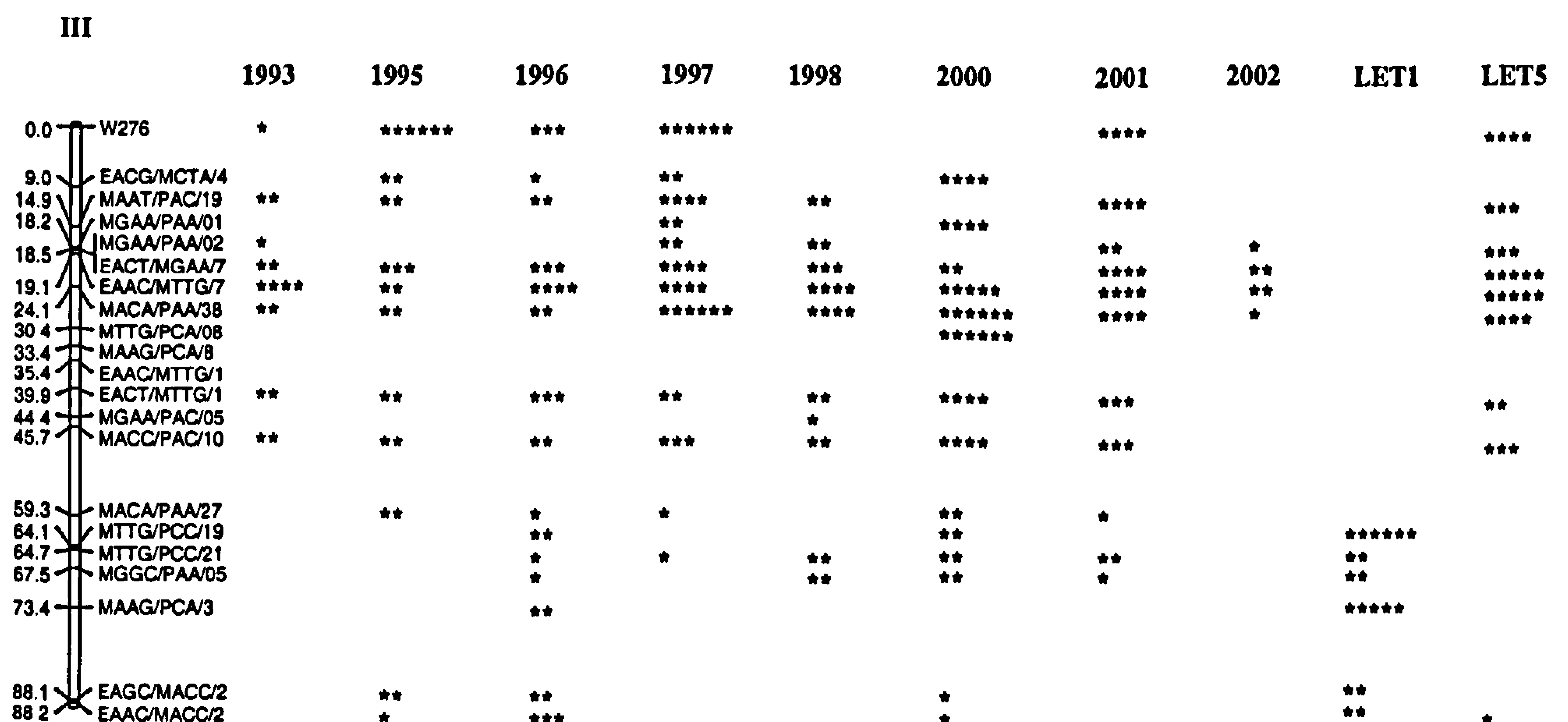
LG	RR-93	RR-95	RR-96	RR-97	RR-98	RR-00	RR-01	RR-02	DM	BF-00	LET1	LET5
III	****	***	****	*****	****	******	****	**				*****
												2.88 (19.6)
V		***	***	****		***	****	****				***
IX											****	
											2.03 (11.9)	
XII					***							
XVIII								***				**
												1.46 (12.4)
F2												
											1.39 (9.5)	

6.3.5.4 The K3 consensus data set

Results of QTL mapping based on the K3 consensus data set are summarised in Table 6.14 for yield-related traits and in Table 6.15 for pest and disease traits. Using IM, one significant QTL for Shts-92 was observed on linkage group I. In addition, several suggestive QTL were identified on linkage groups I (%dry-03, MxHt-03, MxDia-03 and Shts-91), III (Shts-01 and FW-92), VIII (Shts-03), XIII (MxHT-01, MxHt-03, MnDia-01 and MnDia-03), XV (%dry-03 and Shts-91), XVI (MnDia-03), XVII (MxHt-01), XIX (Shts-01 and Shts-91), XX (MxDia-01), Af (MnHt-03 and Shts-91), B (FW-93) and D (MxDia-01). Of these, nine were linked to markers which significant by KW analysis ($P < 0.005$). Furthermore, ten of the suggestive QTL were not detected in analysis based on the parental maps.

As in the parental maps, consistent evidence for QTL on linkage group III influencing field-based rust resistance in all years was attained, for both Swedish and UK data, as shown in Figure 6.10. Moreover, upon analysis of the laboratory rust data, both KW and IM analysis confirmed the likely presence of QTL influencing resistance to both the LET1 and LET5 on this group.

Figure 6.10. Results of Kruskal-Wallis analysis for rust traits on K3 consensus linkage group III. The significance of each marker-trait association is indicated by *asterisks* (*: $P < 0.1$, **: $P < 0.05$, ***: $P < 0.01$, ****: $P < 0.005$, *****: $P < 0.001$, ******: $P < 0.0005$, *******, $P < 0.0001$). Marker nomenclature is as described in Chapter 3, Figure 3.8.



Additional consistent marker associations with field-based rust resistance, that were not revealed in the parental analyses, were detected on several other linkage groups, e.g. markers on group XX showed association with 1996, 1997, 2000 and 2001 rust data. Furthermore, a suggestive QTL affecting resistance to LET5, which was not detected in the parental analyses, was identified by IM analysis on this group.

For the beetle feeding experimental data set, a further suggestive QTL on linkage groups VIII, which was not identified upon examination of either parental data set, was identified. However, no markers on this group were significant for this trait by KW analysis.

Table 6.14. Summarised results of QTL analyses based on growth-related traits and the K3 consensus data (SW980084). For explanation see text.

LG	FW-01	FW-03	DW-01	%dry-01	%dry-03	MxHt-01	MxHt-03	MnHt-03	Shts-01	Shts-03	MxDia-01	MxDia-03	MnDia-01	MnDia-03	FW-92	FW-93	Shts-91
I		*****			**** 3.97 (33.8)		**** 3.54 (27.8)					****** 4.19 (31.5)		****	****** **5.2 (40.5)**		**** 3.56 (23.9)
II	****		****	***	***	****							***				
III					**** near				** 3.7 (53.2)		*****		****		** 3.57 (28.3)	****	
IV				*****							***	***			****		
V											***				***		
VI						***				***							
VII										***							***
VIII										****** 3.26 (40.7)					***	*****	
IX							***										
XII				****													
XIII						*** 3.99(36.2)		******* 3.31 (25.6)					** 2.6 (30.5)	****** 4.23 (36.0)			
XV					**** 2.73 (18.1)												******* 4.1 (27.3)

Table 6.14 continued.

LG	FW-01	FW-03	DW-01	%dry-01	%dry-03	MxHt-01	MxHt-03	MnHt-03	Shts-01	Shts-03	MxDia-01	MxDia-03	MnDia-01	MnDia-03	FW-92	FW-93	Shts-91
XVI			**											** 2.45 (19.9)			
XVII						** 2.73 (29.5)											
XIX									none 2.22 (16.0)								* 2.37 (17.6)
XX											** 2.33 (15.7)						
AF						***		*** 1.97 (15.8)									** 1.73 (11.7)
B																*** 1.45 (9.9)	
D												** 1.55 (10.6)					

Table 6.15. Summarised results of QTL analyses based on pest and disease traits and the K3 consensus data. For explanation see text.

LG	RR-93	RR-95	RR-96	RR-97	RR-98	Rust'00	Rust'01	Rust'02	LET1	LET5	BF'00	Dmarg
I		***	****		***	***			****			
II		***										
III	****	******	****	******	****	******	****	**	****** **5.52 (45.3)**	***** 3.60 (23.8)		
IV				****			***	***		***		
V		***	***	***			****	****		***		
VI			****				****			***		
VIII											3.08 (54.9)	
IX				***					****			
X		***		***			***					
XII					***							
XIII		****	***				***					
XVII								***				
XX			***	****		****	****			**** 3.23 (24.2)		
XXI												*******
Am									**** 2.01 (13.5)			
D				****								
F2									** 1.39 (9.5)			

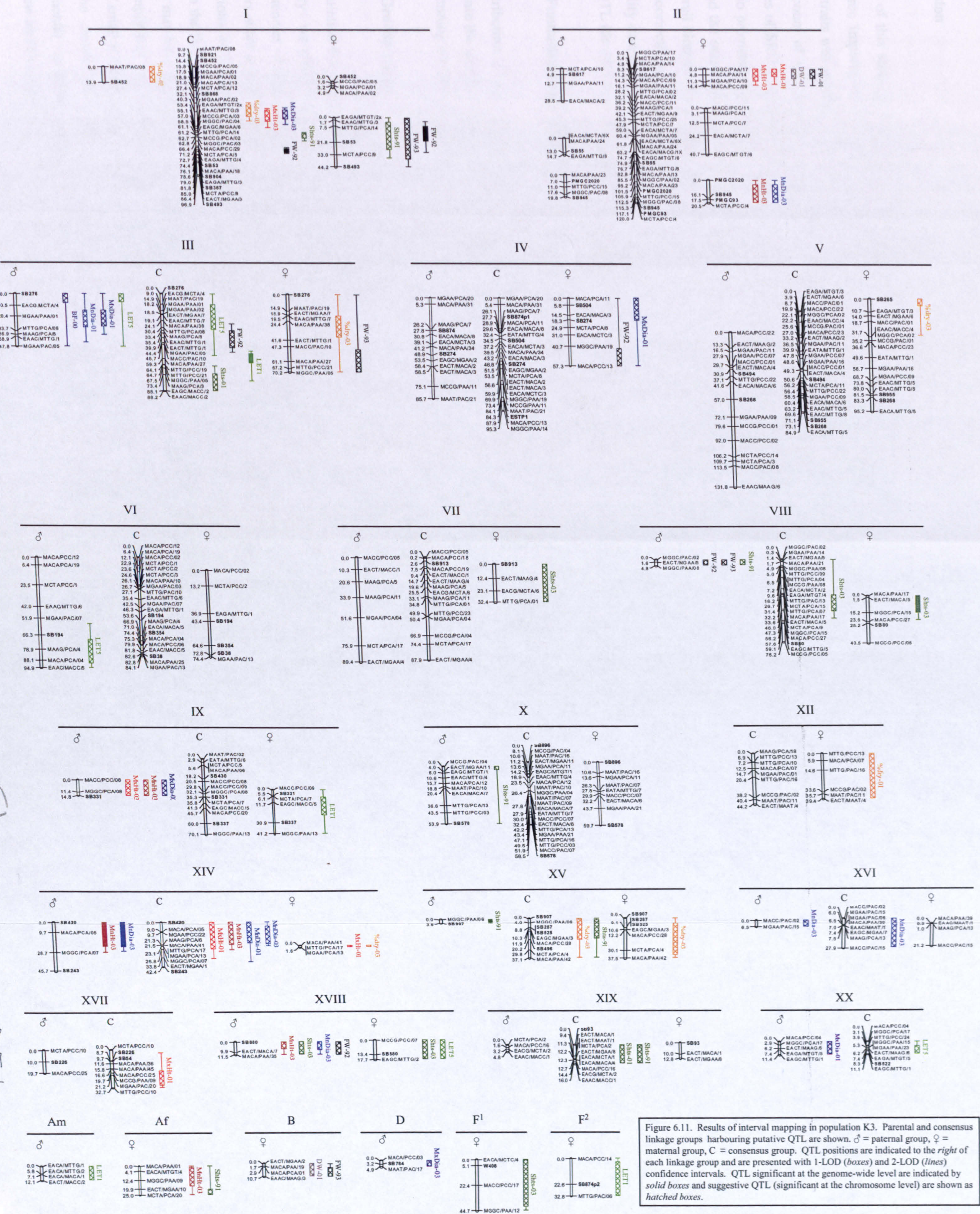


Figure 6.11. Results of interval mapping in population K3. Parental and consensus linkage groups harbouring putative QTL are shown. ♂ = paternal group, ♀ = maternal group, C = consensus group. QTL positions are indicated to the *right* of each linkage group and are presented with 1-LOD (*boxes*) and 2-LOD (*lines*) confidence intervals. QTL significant at the genome-wide level are indicated by *solid boxes* and suggestive QTL (significant at the chromosome level) are shown as *hatched boxes*.

Discussion

Results of this study demonstrated that the identification of QTL affecting traits of agronomic importance is possible in *Salix*. Although detection of QTL affecting growth traits was recently reported by Tsarouhas *et al.* (2001), this thesis provides the first account of QTL mapping results for resistance to the major disease and pest pressures of SRC plantations. Phenotypic assessments of various growth traits in the field also provided some insight into the nature of the genetic control of the different traits and the correlation both among the different traits and between individual traits and overall biomass yield. A limitation of the study lay in the finding that the parents were incorrectly assigned and thus trait data from parents could not be used for heritability studies in this project. In this chapter, aspects of the phenotypic traits and of the QTL identification are discussed.

6.4.1 Fundamental characteristics of the traits assessed

The distribution of trait values within the mapping populations provided an important insight into the genetic basis of the agronomic traits under study, with particular regard to determining the qualitative and/or quantitative nature of each trait.

6.4.1.1 Growth traits

For population K8, examination of phenotypic distributions of all growth-related traits (fresh/dry weights, heights and diameters) following adjustment for spatial inconsistencies within the trial site suggested that genes underlying each of these characters were segregating in this population. Furthermore, as the majority of trait distributions were more or less normally distributed (with the exception of number of shoots in the first year of growth only) it was possible to conclude that these characters were, as may have been expected, inherited via polygenic modes of inheritance. This could imply that the traits were conventional quantitative traits, for which several genes of small effect underlie the phenotype. Alternatively, these traits may have been under the control of fewer genes, each having a relatively large effect, with environmental variation masking the detection of any discrete phenotype classes (Bradshaw and Stettler, 1995).

Mapping studies involving growth traits in poplar (Bradshaw and Stettler, 1995) and willow (Tsarouhas *et al.*, 2002) indicated that the latter scenario may have been applicable in the populations analysed. In both studies, QTL with large effects on phenotype were identified, e.g. in the poplar study 44.7% of the genetic variance in stem volume after two years of growth was attributed to just two QTL; in the willow study, 32% of the phenotypic variance for height growth in a single year was explained by two QTL. In both cases, the question of whether these QTL represented many tightly linked genes, each having a small effect on phenotype, remained unanswered.

For the K3 population, growth trait distributions similar to those for K8 were observed although no adjustments for environmental inconsistencies were performed for this population. However, given that this population comprises a considerably smaller area of the field trial in comparison to K8, it was unlikely that comparable levels of spatial heterogeneity to those recorded across K8 were present in K3. Therefore, for this population the trait distributions were likely to be more representative of variation in the true underlying phenotype values.

6.4.1.2 Rust resistance

In the K8 population the distribution of field-based rust scores deviated from normality in all years tested, with a shift towards a more resistant phenotype observed in all assessment years. The overall distributions suggested that major genes exerting a large effect on the resistant phenotype, as well as several genes of more modest effect, may be segregating within this population.

In 2002, the distribution of rust scores within K8 appeared different to those of previous years, with a larger proportion of the progeny displaying a more susceptible phenotype. While this may have been a consequence of slight differences in the scoring over the years, this could also be explained by altered virulences within the rust population for that season, or perhaps more simply, by more favourable growth conditions for rust in that year.

Over all three assessments, the distributions of rust scores for population K3 were markedly different to those observed in K8. This was most striking in the observation that no K3 individuals displayed complete resistance, although variations in the level

of susceptibility were observed. This indicated that genes conferring an incompatible phenotype (no infection) were not segregating in this population. Furthermore, the near normal distribution of trait scores suggested that, in this population, variation in rust resistance was of a more quantitative nature.

This observation that, within the mapping populations, rust resistance may be both qualitative and quantitative was in agreement with results from a recent study in which inoculation experiments were used to investigate the inheritance of rust resistance in several willow hybrid populations (Pei *et al.*, 2002). Results of these experiments indicated that compatibility to rust was under the control of major genes, with the degree of resistance affected by several genes of additive effect. Similarly in poplar, both qualitative and quantitative resistances to *Melampsora.* rusts have been reported (Lefèvre *et al.*, 1994.)

For the K8 population, distributions obtained for field-based rust scores were in general agreement with those obtained in laboratory-based inoculation tests, with both pathotypes used, i.e. there was a general shift towards resistance accompanied by variation in the range of susceptibilities. It should be noted however, that levels of susceptibility may be slightly increased in laboratory tests in comparison to those observed in the field (Pei *ei al.*, 1996). For population K3, as with the distribution of field scores, the inoculation tests were indicative of quantitative variation for this trait.

6.4.2 Correlations between characters in K8

The examination of correlations between field-assessed traits in K8 served to answer basic questions regarding the importance of both disease and growth components in relation to overall yield. These analyses also served to provide insight in to fundamental breeding issues such as which traits should be selected for within a breeding programme. The ability to examine correlations over a period of time that incorporated establishment, regeneration after harvest and second year growth following harvest, also provided information on the importance of different traits at different stages within the SRC cycle. Of more direct relevance to the current study, the results of correlations between growth components provided a greater opportunity for genetic dissection of broadly descriptive traits such as overall yield via comparison

with QTL. However, it is important to consider that these correlations were true of the population studied and may not be universally applicable to other biomass species, Evidence is now emerging which suggests that different willow species may have different strategies for accumulating biomass (G. Taylor, pers, comm.).

Both height and diameter traits have been identified as highly correlated in biomass willow, both in relation to one another (Tsarouhas *et* al, 2002) and also in relation to overall yield (Rönnberg-Wästljung and Gullberg, 1999). This was in agreement with results obtained in this study in which diameter and height measurements were generally highly correlated with one another and also with yield, both within and across growing seasons. This result suggested that growth patterns for individual trees were consistent over time.

However, in the establishment year, correlations between diameter measurements (MxHt-01 and MnHt-01) with overall yield (FW-01 and DW-01) were weak in comparison to maximum stem height with yield, suggesting that maximum height may be a more important indicator of yield than diameter in early growth stages. The observation that establishment year diameter measurements (2001) were more highly correlated with final harvest weights (2003) in comparison to establishment year heights was also interesting. This result may suggest that selection based on early diameter measurements, as opposed to early heights, may be more indicative of later yields.

The detection of a non-linear relationship between yield and height was also noteworthy. One possible explanation for the exponential increase in yield with increasing height might be provided when considering competition, i.e. individuals that become notably taller than the surrounding canopy provided by their neighbours may be more successful in intercepting light for photosynthesis and thus, are able to accumulate biomass more efficiently. A reflection of this phenomenon in the correlations between maximum diameter (highly correlated with height) and yield may, therefore, be expected

The relatively poor correlations observed for number of shoots per stool and other growth traits, with the exception of harvest weights for which weak to moderate correlations were observed, generally concurred with results of a previous study

examining the effects of several breeding characters on biomass production (Rönnberg-Wästljung and Gullberg, 1999). The weak negative correlations between number of shoots and mean stem diameter observed here in later years were not unexpected given the observation that several K8 progeny with large stem numbers produced many short, narrow shoots in addition to the primary stem(s) (S.J. Hanley, personal observation). This would have resulted in a reduction in the recorded mean stem diameter for each stool as stem number increased. This scenario highlights a bottleneck in willow breeding schemes concerning the choice of the most informative growth characters as a basis for yield selections. The results of this study suggest that assessments of maximum diameter (and also maximum height) may be more informative in predicting yield than assessments based on corresponding mean values derived from all stems per stool.

In terms of overall yield, no evidence for a correlation between rust resistance and performance was identified. This was also the case for all comparisons between rust and growth components, both within and across years. These results, somewhat surprisingly (Section 1.2.3.1.1), suggested that disease levels within the mapping population were not influencing any component of growth to a large degree. This may suggest that the willow genotypes comprising this population display sufficient tolerance to rust infection so that yield remained unaffected. The generally high correlations between rust levels across years suggested that genetic factors were playing a consistent key role in resistance.

6.4.3 QTL analysis: Approach

Separate analyses based on two contrasting populations (K3 and K8), in terms of number of progeny, allowed comparison of the power of QTL detection in small and relatively large *Salix* mapping populations. Furthermore, by including both populations in the QTL mapping study, an opportunity for detection of QTL that were consistent across different pedigrees was afforded. It should be noted however, that the K3 mapping population is related to the K8 population, in that the male parent of K3 (SW870084 var. Orm) is also a parent of the hybrid variety Björn (SW910006) - a great grandparent of the K8 family.

Two complementary QTL mapping methodologies, KW analysis and IM, were used in this study where applicable, in response to suggestions that results obtained by different approaches are more likely to be real and reproducible (Marques et al., 1999). KW analysis is particularly useful for initial marker scans prior to the availability of a linkage, as highlighted in this study by the early detection of the linkage of microsatellite marker SB226 to field-based resistance to rust resistance (Section 5.1.1). Furthermore, this approach is useful for authentication of results obtained from other methodologies (Kearsey and Farquhar, 1998), is more robust with regard to deviations from normality and is the most suitable methodology for analysis of traits that are measured on a categorical, non-continuous scale (Van Ooijen *et al.*, 2002). However, single marker analyses such as this do not take account of all information provided by genotype and trait and linkage data (Liu, 1997). For this reason, interval mapping was also employed (Lander and Botstein, 1989) in this study, to make full use of all available data. The more advanced technique of MQM mapping (Section 1.4.3.1.1) was not implemented here as it is computationally intensive and therefore, could not be included within the timeframe of this study.

In any QTL experiment the question of what significance thresholds should be used to declare the presence of a QTL is an important one. If significance thresholds are set too high, then several QTL may go undetected. Conversely, if thresholds are too low, clearly there is the risk that QTL will be declared when they do not exist. Therefore, for KW analysis, a stringent significance level ($P < 0.005$) was used, corresponding to an overall significance level of around 0.05 (Van Ooijen *et al.*, 2002). The problem of statistical significance in more developed QTL methodologies such as interval and MQM mapping has been addressed in several articles (Lander and Botstein, 1989; Van Ooijen, 1992; Feingold *et al.*, 1993; Churchill and Doerge 1994; Rebaï *et al.*, 1994; Doerge and Churchill, 1996, Doerge and Rebaï, 1996 and Dupuis and Siegmund, 1999) although all of the presented solutions have associated drawbacks, in that they are either computationally expensive, or often involve complex mathematical formulae that are difficult to implement in practise (Van Ooijen, 1999).

For this study, the permutation test of Churchill and Doerge (1994) was used to determine the significance of LOD scores resulting in IM analyses, as it provides a statistically sound method for estimating threshold values, it is robust to departure from standard assumptions (e.g. normality) and is calculated using the data set at hand.

Furthermore, although computationally intensive, it can easily be implemented within the MapQTL® 4.0 software package. As described in the Van Ooijen (1999) two significance thresholds can be considered: the chromosome-wide significance threshold, for which each chromosome (or linkage group) is considered as a separate experiment, or the genome-wide significance threshold which takes into account all of the available marker and trait data. In both cases thresholds are established using a 5% significance level. In this study, as in others (e.g. Atienza *et al.*, 2003), QTL for which associated LOD profiles reach the genome-wide threshold are termed *significant* QTL, while those that exceed the chromosome-wide significance threshold are termed *suggestive* QTL. The use of such nomenclature provide a method by which QTl that are not highly significant, but do exhibit some degree of association with a trait can be reported.

6.4.3.1 QTL analysis : K8

6.4.3.1.1 General considerations

On interpretation of QTL mapping results based on the K8 population, it is important to consider a number of factors potentially affecting the outcome of this study. First, the limited map coverage provided by currently available markers (Sections 5.4.5 and 5.4.6) means that large regions of the genome are not yet represented. Therefore, it may be that only a fraction of the true QTL governing each trait were identified here. Clearly, this reduces the level of characterisation possible for each trait in terms of the true number of underlying loci.

Second, the limited amount of marker information available meant that several consensus map regions were not represented by both paternally- and maternally-derived markers, thus QTL mapping based on the consensus data set was problematic. Although this type of analysis was performed in a preliminary attempt to identify map regions linked to traits in regions comprising markers heterozygous in both parents, analyses based on parental maps should be considered to be more robust in this study.

Third, the genotype data used in the consensus is based on segregation detected with either co-dominant (microsatellites) or dominant marker (AFLP) systems, which, by their nature, can differ in the amount of information they provide (Maliepaard *et al.*,

1997). Although this phenomenon can be largely accounted for in QTL detection by IM analysis based on the all markers approach (where information from neighbouring markers is utilised), results of KW analysis based on single markers must be interpreted bearing this in mind, as the power of each individual test depends on the associated degrees of freedom (Van Ooijen *et al*, 1999). Therefore the more informative co-dominant markers may show a stronger association with a trait than their dominant counterparts, leading to inaccurate predictions of the most likely QTL position. Clearly, this consideration does not apply to analyses based on parental data sets for which all markers are equally informative

Finally, due to time constraints, the data set used for QTL mapping contained several missing genotypes for a number of markers (Section 5.3.5). These missing genotypes would have decreased the power of the experiment by effectively reducing the sample size for individual tests. In future, the resolution of these missing genotypes would improve the accuracy and power of the experiment.

6.4.3.1.2 QTL detection : K8

A large number of potential QTL were detected in this study. Indeed, putative QTL for all traits initially targeted were identified, suggesting that K8 was a good choice of mapping population to underpin this study (Chapter 3).

For growth traits, the magnitude of effect of each QTL was generally low in comparison to those reported in other studies of willow (Tsarouhas *et al.*, 2002) and poplar (Bradshaw and Stettler, 1995). However, the former study was based on a relatively small mapping population (n=94), thus, as acknowledged by the authors, estimations of QTL effect may have been exaggerated (Beavis, 1994). However, one exception was identified in the nursery height QTL on linkage group K, for which the estimation of percentage variance explained was 36%. The possibility remains that QTL of more marked effect are segregating in this population but were not detected as they are located in regions that are not yet represented on the linkage map. Furthermore, several linkage groups on which QTL were likely were only represented by markers from a single parent, e.g. groups P (paternal markers only) and T (maternal markers only). Hence, it is possible that QTL identified in one parent may also be segregating in the other but have, as yet, gone undetected. An additional point to

consider stemmed from the fact that the LOD profiles of several QTL were increasing towards the ends of linkage groups. In such cases, it is likely that the most likely QTL positions have not yet been identified due to the limited marker coverage. Therefore, estimations regarding magnitude of effect (and also significance) will be inaccurate in these instances.

The identification of multiple significant QTL for highly correlated growth traits on group P was striking. Here, QTL were identified for almost all growth traits across several years. In a case such as this, evidence for the validity of the QTL detected is convincing. Whether the QTL identified for several traits on this linkage group represent many different clustered genes, or a single gene with pleiotrophic effect, remains indeterminate. In conjunction with data provided by correlations, the overlapping positions of QTL identified for yield components and overall yield measures may suggest that these traits share common genetic components. However, the most likely position of QTL involved in number of shoots in the second and third year of growth differed slightly on group P compared to QTL for other yield components (Figure 6.9). This trait was not highly correlated with any other yield-components across the assessment years, suggesting that on this linkage group, QTL underlying this trait might be distinct from those governing other yield components. The co-location of height and diameter QTL has previously been reported in willow by Tsarouhas *et al.* (2002). Similarly in poplar, clustering of QTL for stem basal area and number of sylleptic branches has been described (Bradshaw and Stettler, 1995). Co-location of multiple yield-related traits both within and between years, was observed on several linkage groups other than P, most notably on groups A an C, but also with lesser consistency on others.

The identification of QTL that were consistent over time is in contrast to the study of Tsarouhas *et al.* (2002), where no such consistency was apparent. Although there were obvious differences between this study and the one described here (different environments, different planting regimes, different pedigrees, etc.), these contrasting results may be a reflection of the increased power of QTL detection, in terms of population size, afforded by the larger K8 population. However, it is important to note that several QTL were also identified in the current study which were specific to a given year or growth stage. As highlighted by Tsarouhas *et al.* (2002), this may not be surprising for willows grown as SRC due to the drastic changes in shoot-root growth

relationships over the rotation cycle. Interestingly, linkage groups K and T showed some evidence of containing QTL that may be specific to an early stage of growth. For linkage group K, of all of the putative QTL detected, those for yield (FW-01 and DW-01) and maximum height (MxHT-01) were most significant in the establishment year. Furthermore, a significant QTL for nursery height was detected on this group. Similarly for group T, significant and suggestive QTL were only detected for yield-related traits in early stages of growth, i.e. the establishment year and in the nursery.

The inclusion of nursery height (NH) in analysis provided a means for QTL validation as it presented QTL results for a growth-related trait that was not assessed in the field trial as for other traits, but in a contrasting environment. In several instances, QTL governing this trait co-located to linkage groups that were implicated for growth traits upon analysis of field-based data, e.g. groups A, K and P. This suggested that the QTL identified using data from both environments were real. Furthermore, this approach meant that QTL that were putatively either environment- or growth stage-specific could be identified. For example, a suggestive QTL for NH was identified on linkage groups M and Z. However, with the exception of percentage dry matter on Z, no other QTL growth traits were detected on either group. Therefore, this QTL may be particularly important for growth at an early stage.

The ability to map QTL for overall yield was demonstrated in this study. In willow, the possibility of selecting for an ideotype for overall biomass yield has been discussed however, a large number of different traits could be important to consider and in order to achieve this goal it will be necessary to first establish which characteristics are important for high biomass yield. The proportion of variation that is genetically determined in all of these traits and the proportion of the correlation between the traits that is genotypic, should then be established (Viherä-Aarnio, 1988). This is not a trivial task and although some progress was made towards it here, a thorough investigation was outside the scope of the present work. As the number of QTL that can be identified for a given trait in a single study is generally limited (Hyne and Kearsey, 1995), the presumed complex nature of the trait may restrict the likelihood of identifying QTL for overall yield. For this reason, it may be beneficial to focus QTL mapping efforts on more simple traits that are identified as playing an important role in determining yield. Such traits are referred to as yield components in the current study. Data from previous studies and correlations identified here suggested that yield

components such as stem height, stem diameter, number of shoots per stool may all play a role in determining overall yield and may therefore be suitable targets for QTL mapping. In this study, the significances associated with QTL identified for yield components were generally higher than those for overall yield assessments. However mapping QTL for overall yield was also achieved, indicating that both strategies are viable in biomass willow and that a few components may have a large effect on overall biomass yield.

The identification of QTL for percentage dry matter was encouraging as this trait may particularly important in biomass willows considering their end-use as a fuel source. However, there was little consistency between QTL positions for this trait between the two years in which assessments were performed. This may be a consequence of the two different growth stages involved, i.e. following the establishment season (2001) and also in the second year of growth (2003). One notable exception, however, was the identification of highly significant QTL for this trait on linkage group B in both years assessed.

For field-based rust resistance, the identification of highly significant associations between markers on linkage group I may be suggestive of a major gene (or genes) on this linkage group. This may also be applicable to the QTL consistently identified on linkage group B. However, as expected from trait distributions and previously reported studies on rust resistance in willow (Pei *et al.*, 2002), the identification of additional QTL on different linkage groups was commensurate with the expectation that rust resistance is also of a quantitative nature. The results of QTL analysis based on rust data from the field were reflected on analysis of data provided by inoculation tests, particularly for the LET1 inoculum. Here a significant QTL was identified on linkage group I in the resistant parent (K477-R13). Additional evidence of QTL was observed for linkage groups A and E (also positive with field data) using IM analysis. For the LET5 pathotype, evidence for associations were also detected on linkage groups A and I by KW analysis, although in both cases significance thresholds were not achieved. This may reflect the limited power of QTL detection associated with the inoculation test data due to the relatively small number of individuals for which trait data was generated. In future, it may be possible to validate these QTL if the inoculation experiment is repeated to include a greater number of progeny.

Interestingly, a number of likely QTL positions for field-based resistance to rust co-located with QTL identified for percentage dry matter content. This was most notable for linkage group B, but examples of this were also observed on linkage groups A, K and F. However, little evidence for this was obtained for linkage group I, which showed the strongest association with rust resistance. This may be an example of where QTL analysis can shed light on the causal nature of observed trait variation, as the suggestion here is that susceptibility to rust may have influenced variation observed for percentage dry matter content. This may explain why a number of outliers in the trait distributions, with higher dry matter content, were observed. These progeny may have contained dead, dry shoots as a result of severe rust infection, and would therefore have shown a shift towards a higher percentage dry matter content. To investigate this further in future, it may be useful to discard these outliers and repeat the QTL analysis. Examination of correlations between rust levels and percentage dry matter content (not performed within this study) may also prove informative.

Despite the expectation that identification of any differences in palatability to willow beetles amongst the K8 progeny may be difficult (Section 1.2.3.2.1), some evidence for the presence of underlying QTL was detected. Furthermore, as with the rust inoculation tests, the power of QTL detection was lower than for other field-based assessments due to the smaller number of progeny for which trait data was determined. One putative QTL in particular looked interesting. This suggestive QTL, mapped to linkage group S, was detected upon analysis of data generated in both the 2000 and 2001 laboratory-based feeding experiments. Although genome-wide significance was not achieved in either instance by IM analysis, the association was significant at the chromosome level and also by KW analysis. The fact that this linkage group was highlighted in two completely independent assessments may be indicative of its validity although, clearly, further testing will be required. The inconsistency in positions of additional putative QTL identified between the two independent experiments may reflect the different nature of the leaf material used in each assay. In the first experiment, leaf material was collected from regenerated rootstocks grown in a gauzehouse. Consequently leaves were less tough than those obtained from the field for use in the second experiment (S. Hanley, personal observation). Thus, QTL inconsistency may also be explained here in terms of differences in environment. Alternatively, as the beetles used in the two tests were collected from different sites,

there may have been slight variation in feeding preferences associated with the two beetle populations. Evidence for geographically-based differences in feeding preference has been reported by Green *et al.* (2001) and Karp and Peacock (2003).

6.4.3.2 QTL analysis : K3

6.4.3.2.1 General considerations

As for K8, several factors should be considered when interpreting QTL results in K3. First, the size of the mapping population (n = 64) is relatively small for QTL analysis (Beavis, 1994; Young, 1999). Therefore, only QTL of relatively large effect are likely to be detected in this family (Hanley *et al.*, 2002). It is also likely that the percentage of the phenotypic variance explained by resulting QTL will be exaggerated due to the small population size. Second, the genome coverage provided by the map is not complete (Hanley *et al.*, 2002), thus any QTL in regions that are not represented will have gone undetected. Furthermore, for the consensus map, linkage groups vary in the density of markers derived from each parent, thus QTL analysis based on the consensus data may not be comprehensive. Finally, both dominant and co-dominant markers were used to construct the map. The implications of this approach are discussed in Section 6.4.3.1.1.

6.4.3.2.2 QTL detection : K3

In general, less QTL were detected in this population in comparison to K8. This is most probably due to the small population size but may also be a reflection of the fact that the K3 pedigree comprised only pure *S. viminalis* individuals. The latter explanation may be less likely given the evidence from the willow QTL study of Tsarouhas *et al.* (2002) in which a greater number of growth trait QTL were detected in the pure *S. viminalis* parent of the mapping population in comparison to the hybrid *S. viminalis* x *S. schwerinii* parent Björn. This result indicated that there is high genetic variation for growth traits within *S. viminalis*, suggesting that QTL mapping in an intra-specific cross is viable.

In terms of consistency of QTL detection across years and across potentially correlated traits, this was less evident in K3 compared to analyses of the larger K8 population.

While these results may have been truly representative, it is probable that the small population size prevented the detection of significant (or suggestive) QTL across assessments, as in K8, the significance of QTL across years was found to be variable. This may have also have been a limiting factor in the study of Tsarouhas *et* al. (2002). Despite these limitations, some linkage groups showed a degree of consistency in terms of their association with related traits. For example, evidence for QTL affecting MxHt-03, MxDia-03, FW-92, FW-93 and Shts-91 was detected on linkage group I. On linkage group II, there was evidence to support the presence of QTL affecting FW-01, DW-0, MxHt-01, MxHt-03, MxDia-01 and MnDia-01. Once more however, the significance of each of these putative QTL was found to be variable, especially on linkage group II, where QTL for establishment year traits were more convincing than those identified in later years. This may suggest that these QTL play a greater role in growth at early developmental stages.

Whilst QTL analysis of growth traits using both the Swedish data (see Table 6.1) and the UK data sets failed to detect any consistent QTL positions for any single trait, there was evidence to support the presence of QTL for potentially related traits on K3 linkage group I. Here, putative QTL were identified for maximum height and diameter in the final UK assessment and for fresh weight in 1992 and 1993 in Sweden. Although not examined directly in this population, in K8, heights, diameters and fresh weights were highly correlated (Section 6.3.3). If these traits are correlated in K3, and it is reasonable to assume that they are, then the limited consistency of QTL across environments and time may be more of a reflection of the power of QTL detection power in a small population, rather than a representative description of the presence or absence of QTL. However, the possibility that QTL specific to either the Swedish or UK environment were detected must be considered. Such differences may be expected given the different climates of these two countries.

The detection of consistent putative field-based rust resistance QTL across several years suggested the validity of results. Moreover, several linkage groups displayed an association with this trait in both environments examined (Sweden and the UK). The detection of putative QTL on a particular linkage group in some years but not others may be consequence of changing pathogen virulences across seasons. The results of QTL analysis of resistance data generated in the inoculation tests were also in general agreement with field data in terms of the linkage groups implicated. Strong evidence

for the presence of a QTL affecting resistance to both the LET1 and LET5 pathotypes was detected on linkage group III, for which the strongest associations with field-based resistance were also detected. Furthermore, for the homeologous linkage group in K8 (Group B; Section 5.3.11), significant evidence for a consistent QTL implicated in rust resistance was also identified. This result substantiates the validity of these respective QTL in both populations. Interestingly, no significant association between rust resistance on linkage group XVII in K3 were detected. The homeologous linkage group in K8 was group I, for which the most highly significant QTL for rust resistance were recorded in this population.

Although not conclusive, comparisons of QTL detected using the inoculation test data and those identified with field-based data were interesting. For example, considering the field assessment data alone, where a putative association between linkage group I was identified in a given year, there was no evidence of a significant (or near significant) association between resistance and linkage group IV, i.e. for RR-95, RR-96, RR-98 and RR-00 in the same year. In the inoculation tests, the putative QTL for resistance on linkage group I was only identified when LET1 was used as the inoculum. Furthermore, there was no convincing evidence to suggest that there was a QTL involved in resistance to this pathotype on linkage group IV. In contrast, using LET5, these results were reversed, i.e. there was no evidence for a QTL on linkage group I but there was a suggestion of a QTL on group IV. This may suggest that the prevalent rust pathotype in the field in 1995, 1996, 1998 and 2000 shared a virulence pattern more similar to LET1 than to LET5, and that the putative QTL on linkage group I is of particular importance for resistance to LET1-like pathotypes

For beetle resistance, results of QTL analysis suggested that there was only slight evidence for the presence of QTL affecting this trait. The putative QTL identified on linkage group III on analysis of paternal data set was significant by both KW and IM analysis at the chromosomal level and may, therefore, be of interest. However, the QTL that was identified by IM (chromosomal significance only) on linkage group VIII on analysis of the consensus data set, showed no signs of an association by KW analysis. Therefore, the reliability of this QTL may be questionable. The failure to detect any strong associations for this trait is most likely a consequence of the pure *S. viminalis* pedigree, within which there may be little difference in beetle feeding preferences (Section 1.2.3.2.1).

Although not initially a target trait in this study, the identification of highly significant associations between markers on linkage group XXI and resistance to the leaf roll galling midge (*D. marginemtorquens*) is of potential interest to willow breeding programmes. Although this insect has not yet been reported as causing significant damage to willow crops in the UK, it is considered an important pest in Sweden (Glynn, 1996) and resistance is taken into account for selection within breeding programmes (Larsson, 1996).

6.5 Chapter summary

- This chapter describes the use of previously constructed linkage maps (Chapters 3 and 5) and phenotypic data generated in several trait assessments in QTL analysis aimed at the identification of genomic regions underlying variation in traits of agronomic importance. In addition, more fundamental issues regarding relationships between various traits were addressed.
- In addition to trait assessments described previously in Chapter 3, results of further data collections for yield-related traits and also field-based resistance to rust are described.
- A second, more in-depth laboratory-based beetle feeding was performed to provide additional trait values for use in QTL analysis.
- Phenotypic correlations based on all K8 field-based traits were calculated. Several strong correlations between different yield components and between yield components and overall yield were identified. In addition, information regarding correlations between characters over time was obtained.
- Prior to QTL analysis, all K8 field-based trait data was adjusted for any spatial heterogeneity across the trial site using the reference clone (var. Jorr) as a standard.
- Linkage and trait data for both the K3 and K8 mapping populations was used in preliminary QTL analyses based on parental and consensus linkage maps.
- For yield and yield components, several QTL were identified in K8. Several reliable QTL positions for traits assessed at different times were detected. In particular, linkage groups A and P were consistently highlighted as harbouring QTL for several correlated yield and yield-related traits.

- In K3, QTL for yield and yield components were identified, although the consistency with which they could be detected was low in comparison to K8 – this was most likely a consequence of decreased power of QTL detection in K3 afforded by the relatively small number of progeny.
- A major QTL for resistance to rust was detected on K8 linkage group I. A second significant association with this trait was detected on linkage group B. This latter QTL was also evident on K3 linkage group III (homeologous with K8 group B) and was stable across several years and across Swedish and UK environments. Several QTL for field-based resistance were also detected upon analysis of data generated in laboratory-based inoculation experiments.
- No highly significant QTL for resistance to herbivory by the the blue willow beetle (*P. vulgatissima*) were detected, although some suggestive QTL were identified in K8. One such putative QTL was identified on K8 linkage group S following analysis of data generated in both the 2000 and 2001 beetle feeding experiments.
- A significant association between K3 linkage group XXI and resistance to the leaf roll galling midge (*D. marginemtorquens*) was detected.

7. General Discussion and Future Perspectives

As stated at the outset, this thesis describes the results of genetic mapping of agronomic traits in willows grown for biomass production through the application of molecular marker techniques. This chapter provides a general appraisal of the results obtained in this endeavour, in the context of the initial aims and specific objectives of the project, and also discusses implications for future work.

Molecular markers

At the start of this work, DNA-based molecular marker research was limited in willow species in comparison to many other crops, including poplar, with only a handful of publications available at that time. The vast majority of these studies reported investigations of genetic diversity within the genus (Lin *et* al., 1994; Lindegaard and Barker, 1997; Triest *et al.* 1997; Barker *et al.*, 1999) although in some instances these investigations were aimed at resolving genetic relationships within breeding germplasm. In all of these studies, molecular marker analyses were limited to the use of arbitrary marker classes such as RAPDs and AFLPs with the availability of co-dominant marker systems such as RFLPs and microsatellites not reported. Efforts to generate microsatellite markers specifically for use in willow were initiated at LARS with the production of a microsatellite-enriched library (Edwards *et al.*, 1996) and the subsequent development of informative markers during the EU-AIR project (Barker *et al.*, 2003). However, at the time that work described in this thesis commenced, the number of available microsatellite markers remained limited and was not nearly sufficient for use in a mapping study of *Salix*. Further characterisation of this library and subsequent development of a large number of microsatellite markers during this project produced an important resource for future studies in willow, not only in terms of linkage mapping studies, but also for other genetic applications such as diversity studies and pedigree testing. Of the markers tested for mapping in K3 or K8, a significant number were not used in mapping as they failed to segregate within either or both populations. However, while this was a bottleneck for mapping in both pedigrees analysed here, these markers could be informative in other genetic studies of either mapping or natural populations of willows.

The usefulness of microsatellite markers was well demonstrated here, not only in terms of their highly informative (co-dominant) nature in linkage and QTL analyses, but also with regard to the identification of the correct parents of the K8 mapping population. Within linkage analyses, microsatellites were of particular use in that they provided allelic bridges to allow the integration of separate male and female maps. Furthermore, due to their sequence-based nature, they also allowed identification of a number of putatively homeologous linkage groups between the K3 and K8 maps. However, the information provided by this approach was restricted due to the limited number of microsatellites that could be mapped (i.e. that were polymorphic) in both pedigrees. Had more time been available, then microsatellite markers produced in the second phase of development (SB984 onwards; Table 5.4) for mapping in K8 would have been tested for their ability to detect polymorphisms between the K3 parents.

A potential drawback of using microsatellite markers in linkage analysis results from their single locus nature. This can mean that these markers require significant investment of time and resources to be developed and mapped, especially when large numbers of markers are to be mapped on a large number of progeny. However, in the mapping study based on the K8 population, this hurdle was largely overcome through the development of microsatellite multiplexing strategies based on fluorescent detection systems. The implementation of these approaches based on both PCR and loading multiplexes, saved considerable time and money but were only made possible in the latter months of the project when an ABI 3100 Genetic Analyser was acquired. However, in some instances the mapping of microsatellite markers via more conventional radioisotope methods proved invaluable, indicating that the implementation of both technologies may be beneficial in studies such as this.

As may have been expected (Jones *et al.*, 1997) AFLP markers were highly reliable in this study, and were particularly useful for attaining rapid map coverage (Chapters 3 and 5) and in pedigree testing of the candidate crosses for use in this study (Chapter 4). However, in linkage mapping it was necessary to treat AFLP markers with caution if highly distorted segregation ratios were detected, especially if the markers showing distortion did not map to regions in which linked markers also displayed distorted segregation. This situation substantiated one of the reported problems associated with this particular marker technology, in that several bands of equivalent electrophoretic mobility may co-migrate to the same gel position. Clearly, inclusion of such markers

in linkage analyses is problematic, so for this reason, several AFLP markers were discarded from the current study. Despite this minor limitation, AFLP technology proved highly valuable in this study as large numbers of informative markers were generated in relatively short timeframes.

Although not used to any large extent within current study, the efficacy of using ESTP markers was confirmed, providing a potential route to the development of markers based on functionally relevant sequences. The validity of this approach has been reported in an ESTP mapping study of loblolly pine (Temesgen *et al.*, 2001). However, it is important to note that while the sequences from which the two markers examined in this study were derived from a poplar EST sequencing program, the expression of these sequences in willow was not confirmed. This may be investigated in future through the use of reverse transcription PCR (RT-PCR) (Freeman *et* al., 1999) or hybridisations based on RNA (e.g. a Northern blot; Sambrook *et al.*, 1989). However, The ESTP approach demonstrated the potential exploitation of poplar genomic resources (i.e. EST databases) in willow research. This is an important consideration for future willow studies, for which the ever-increasing amount of genetic information generated for poplar may be exploited. This will become even more relevant at such time that the complete poplar (*P. trichocarpa*) genome sequence becomes available (6x draft sequence expected by the end of August 2003; DOE Joint Genome Institute, USA).

The drawback associated with the development of markers based on coding sequences results from the fact that there is a tendency for sequences within coding regions to be more highly conserved than those in non-coding, intergenic areas (Cato *et al.*, 2001). Consequently, the identification of suitable polymorphisms for mapping by conventional marker technologies, which are often based on detection of size polymorphisms, may be difficult. Perhaps the most likely solution to such problems is provided by the interrogation of SNPs, which are the most common form of polymorphism between alleles (Rafalski, 2002). Recent advances in sequencing technologies make the identification and use of SNPs an attractive route towards EST mapping in future.

Given the significant investment required, future mapping of ESTP markers in willow may be most efficient if targeted towards functional candidate genes. A similar

approach has been used in poplar in a study of QTL for bud set and bud flush, where two candidate genes believed to be involved in perception of photoperiod (*PHYB2*) or transduction of abscisic acid response signals (*ABI1B*) co-located with QTL for bud set and bud break (Frewen *et al.*, 2000). While these results did not prove a definitive role for the candidates in the trait, they did provide targets for further validation studies such as statistical association testing (for linkage disequilibrium) in natural populations or genetic transformation studies (Pflieger *et al.*, 2001). As the amount of sequence data available via public databases continues to increase, and the role of greater numbers of gene sequences is elucidated, the mapping of candidate genes becomes an increasingly more attractive approach.

A second application of the candidate gene approach in future willow mapping studies may focus on the development of molecular markers based on Resistance Gene Analogues (RGAs) (Section 1.4.7). Although not included in this thesis, preliminary work is currently underway towards this goal in willow, using degenerate primers designed to the nucleotide binding site (NBS) region of known R-genes (Kanazin *et al.*, 1996) to amplify willow RGAs. Resulting products have been cloned and sequenced and the homology to known R-genes (or RGAs) confirmed. However, initial examination of the resulting sequences suggested that multiple highly homologous, but unique, sequences were obtained using this approach. Consequently, RGA mapping in willow will be complicated and conventional mapping strategies such as those based on restriction digests may not be applicable in many cases. For these reasons the RGA approach was not pursued further as part of this thesis.

The mapping populations

The establishment of the large K8 mapping population to underpin QTL studies described in this thesis represented the development of an important genetic resource to underpin trait analyses in willow. This population represents the largest segregating mapping population currently available for willow. Using this population alone, putative QTL for all traits initially targeted in this study were identified. This highlights the usefulness of this population as an underpinning resource for future QTL investigations. However, this is not to say that other segregating populations will not be required. For example, interspecific crosses between highly palatable species and

those which are highly resistant to willow beetle herbivory may be more suited to identification of genomic regions underlying this trait.

The ease of clonal propagation in willow, via the production of cuttings, means that this population is essentially immortal and could therefore be used as a basis for QTL mapping studies in several different environments and could be made available to the willow research community. Already this populations has been replicated in a field trial at Rothamsted Experimental Station, Herts, UK not only for future validation of QTL identified in this study across two contrasting environments, but also to allow for the identification of any environment-specific QTL.

Trait analyses

Trait assessments performed in this study have provided data used to answer some fundamental questions regarding the basis of some important traits in biomass willow. For example, analysis of segregation of rust resistance scores indicated that this trait is not purely qualitative and is influenced by multiple factors that are likely to differ in the effect they exert. For yield, results indicate that this trait is of a more quantitative nature, in that numerous genes are likely to be of importance. Answers as to exactly how many genes are important are as yet undetermined due to the inherent limitations of QTL analysis (Kearsey and Farquhar, 1998) and the incomplete genome coverage provided by currently available markers (Chapters 3 and 5).

Analysis of phenotypic correlations between characters shed light on the relationships between various distinct traits that may potentially be indicative of biomass yield. For example, maximum stem height and stem diameter were highly correlated with overall yield in the K8 mapping population. Results such as these are informative in that they suggest targets for future molecular study and also may also highlight which traits may be informative in selections within a breeding programme. However, it is important to note that it is possible that several traits not examined in this study may also be of importance. For example, preliminary physiological studies in a high yielding variety (var. Tora) and low yielding willow ("L78183") have highlighted leaf extension rate and leaf cell number as a potentially important trait affecting overall biomass yield (Robinson *et al.*, in press). Such studies may provide additional targets for future QTL mapping and candidate gene analysis.

The linkage maps

The construction of both the K3 and K8 linkage maps represented an important step forward in the molecular genetic analysis of *Salix* by providing fundamental genetic resources to underpin further research. However, as expected, the application of linkage analysis in willow required significant investment of both time and resources. The development and screening of large numbers of markers was required in order to provide an acceptable level of genome coverage. This was a consequence of the relatively large number of chromosomes (n=19) present in this species. Furthermore, both mapping populations were generated from crosses between parents that were themselves full-sibs, thus, the identification of suitable numbers of polymorphic markers, in particular microsatellites, was more difficult than it would have been if the mapping populations had been derived from more genetically diverse parents.

The K3 linkage map was more highly saturated with molecular markers than the K8 map and may, therefore, be an important source of markers for mapping in regions not yet represented on the K8 map, via the conversion of AFLP markers to SCARs. This approach was not implemented in the current study due to time limitations but should be achievable in future. As well as improving the genome coverage for the K8 map, this would also allow for more accurate comparisons between maps, as a greater number of shared markers would be identifiable. However, the mapping resolution power afforded by the K3 population, even if the additional available progeny were included in future analyses (Section 4.1.2), would not be as great as that provided by the larger K8 population. Similarly, the power and accuracy of any QTL studies in the K3 population will not be as high as those based on K8 (Beavis, 1994; Hanley *et al.*, 2002). It may therefore be prudent to concentrate all future linkage mapping and marker development efforts on the larger of the mapping populations.

Despite the preliminary nature of the K8 linkage map as reported in this thesis, the available linkage information served as a basis for early QTL analysis and provided a means of comparison of marker (and QTL) positions, via the microsatellite markers, with the K3 map. However, a significant amount of work remains in order to maximise the information content provided by the K8 map, in that the genome coverage still needs to be improved and the missing genotypes should be determined to promote greater accuracy, both in terms of linkage and QTL analyses.

The availability of a willow linkage map comprising markers derived and mapped in poplar (and vice versa) may be beneficial to future research as comparative mapping will become a possibility. The efficacy of using poplar microsatellite markers in willow was demonstrated here, although the proportion of poplar microsatellites that failed to amplify in willow (~75%) suggests that sufficient diversity is present between these genera to prevent the straightforward exchange of markers. Therefore, comparative mapping may be more successful if based on more highly conserved expressed sequences, i.e. ESTs, in future.

QTL analysis

Despite the limitations outlined in Chapter 6, QTL analyses based on both the K3 and K8 population were successful in so far as genomic regions linked to traits of agronomic importance were identified. As expected, the larger progeny size available for analysis in K8 resulted in greater QTL detection power, as illustrated by the larger number of QTL uncovered in this population compared to K3. However, this may have been due to the segregation of a greater number of QTL in population K8. While this may have been expected to some degree, given that the K8 pedigree comprised both *S viminalis* and *S. schwerinii* species while the K3 population comprised only pure *S. viminalis*, results reported by Tsarhouras *et al.* (2002) suggest that this may not be the case, as a larger number of QTL were detected in the pure *S. viminalis* parent of the mapping population used compared to the hybrid *S. viminalis* x *S. schwerinii* second parent. The identification of putative QTL underlying all traits initially targeted in this study, illustrated the suitability of the K8 population for QTL mapping of important agronomic traits in biomass willow.

Putative QTL were identified for total yield and for all of the components of yield assessed in the K8 mapping populations. In K8 several consistent associations were detected indicating the robustness of the QTL uncovered. Evidence of reliability was afforded by the detection of QTL that were consistent between related yield traits assessed in different environments, e.g. QTL were identified on K8 linkage group K for height measured in the nursery, height measured in the field during the establishment year and overall yield following the first harvest. In addition, the detection of consistent QTL for numerous correlated traits over time was indicative of

the dependability of the QTL detected, e.g. several QTL were identified for related traits on linkage group P across all assessment years. While the co-location of QTL for correlated yield-based traits was previously been reported in willow (Tsarouhas *et al.*, 2002), the consistent detection of QTL across time and different growth stages as shown here, was not. This may have been a consequence of the relatively small mapping population used in their study. Support for this explanation comes fromthat the fact that the QTL mapping results obtained here with the relatively small K3 mapping population were also less consistent than those obtained with the larger K8 progeny.

The identification of QTL for rust resistance that were consistent over both time and environment, i.e. on K8 group I and on K3 and K8 homeologous groups III and B, respectively, suggested that these associations were robust. These highly significant rust resistance QTL may be worthy of further study. However, it is important that a number of points are first addressed. First, missing genotypes for markers in these regions should be resolved to ensure that each marker affords an equivalent amount of information. To a similar end, the co-dominant microsatellite markers should be analysed as dominant markers. More accurate positioning of the QTL may then be possible. Once achieved, nearby markers could be tested in marker-assisted selections.

Preliminary data suggest that microsatellite marker SB226 may be of use in such an approach. To illustrate, when SB226 was used to screen the K477 population in experiments to identify the correct K8 parents (Chapter 5) two alleles were detected (data not shown). The inheritance of one of these alleles in the K8 population was associated with a greater level of rust resistance. In the K477, with the exception of K477-R10, all of the individuals designated as rust resistant carried this allele. However, K477-R10 had a more susceptible rust score in 1996 (Table 4.2) than the other resistant sibs and may, therefore, have been mis-classified. With the exception of K477-S3 (the K8 female parent), all of the remaining designated susceptible individuals lacked this allele. The mean field-based rust score for K477-S3 was at a comparable level to other susceptible sibs (Table 4.2), thus, in this individual a recombination event may have occurred between SB226 and the resistance locus. Alternatively, original rust scores may have been inaccurate for this clone. Unfortunately, susceptibility/resistance to rust in K477-S3 could not be confirmed in

the current study due to the inclusion of the incorrect parents in rust assessments (Section 5.4.1).

Rust resistance within many of the commercial biomass varieties is believed to have originated from the introduction of a highly resistant Russian *S. schwerinii* clone (L79069) into the breeding germplasm (Larrson, 2001). This clone is a parent of the Swedish variety Björn – a great, great grandparent of the K8 mapping population. Therefore, the resistance within K8 may have been inherited from L79069. Preliminary screens have now shown that the SB226 allele putatively associated with resistance in K8 is also present in all of the resistant willows contained in the pedigree, from L79069 through to the K8 population, i.e. L79069, var. Björn (and it's highly resistant full-sib var. Tora), SW930984 and K447-R13. However, with the exception of K477-S3 (susceptibility not confirmed in this study), this allele is not present in any susceptible willows within the pedigree, i.e. var. Orm or var. Astrid. Although these results are only based on a small number of samples, further screening across the breeding material should now be performed to confirm any predictive value of this marker. However, the possibility of testing other markers in the vicinity of the original QTL should be considered, as these may be more tightly linked to the QTL than microsatellite SB226. This may require the conversion of AFLPs into SCAR markers for ease of screening.

If the position of this QTL can be more accurately determined and tightly linked markers are available, this locus may be a suitable candidate for map-based cloning. Cloning of the underlying resistance QTL may subsequently facilitate selections based on the gene itself, which may be more efficient than using linked markers. The small genome sizes of *Salix* species (2C = 0.76-0.98 pg; Thibault, 1998) makes them particularly suited to map-based cloning. However, it will first be necessary to construct a large-insert library (e.g. a BAC library) for this purpose and further work will be required to finely map any important QTL-containing regions.

For resistance to herbivory by the blue willow beetle (*P. vulgatissima*), a number of putative QTL were identified based on data generated in laboratory-based feeding experiments. However, in general, the significance of these associations was poor, although it must be noted that, as with all other traits examined, incomplete genome coverage provided by the markers available may have prevented the identification of

more highly significant QTL. In spite of this limitation, one suggestive QTL identified upon analysis of data from the two independent assessments (on K8 linkage group S) may be worthy of further study, although this may be better achieved in populations derived from parents of more contrasting phenotypes.

In order to improve the power of QTL analysis in subsequent analyses based on K8, it is important that marker coverage on the map is improved. This may help to substantiate the significance of suggestive QTL that were identified in large marker intervals, e.g. on K8 linkage group K, and may also decrease the large confidence intervals associated with many of the QTL detected. Bulked segregant analysis may provide a powerful approach by which markers targeted to potentially interesting map regions can be developed in future. This approach was successfully employed within the current study to map markers in the vicinity of microsatellite marker SB226. Furthermore, the inclusion of the K8 progeny that were not utilised in current analyses may be of potential benefit in terms of increasing mapping resolution, substantiating the significance of QTL identified here and increasing the accuracy with which QTL positions can be determined.

Summary

To conclude, this project has been successful in meeting the overall aims and objectives as set out at the onset of the study. Additional molecular markers were developed and successfully used to generate linkage maps and identify genomic regions underlying traits of agronomic importance in biomass willow. Moreover, putative QTL for all traits initially targeted (resistance to rust, resistance to leaf-eating beetles and yield) were identified. Although still preliminary in some cases, the results presented in this thesis represent a significant step forward in terms of making molecular breeding of SRC willows an obtainable future prospect and provide many avenues for future research into this increasingly important crop.

References

Aarts M, Hekkert B, Holub E, Beynon J, Stiekema W and Pereira A. 1998. Identification of R-gene homologous DNA fragments genetically linked to disease resistance loci in *Arabidopsis thaliana. Molecular Plant-Microbe Interactions* **11**, 251-258.

Acarkan A, Rossberg M, Koch and Schmidt R. 2000. Comparative genome analysis reveals extensive conservation of genome organisation for *Arabidopsis* and *Capsella rubella. Plant Journal* **23**, 55-62.

Aharoni A, Keizer LCP, Bouwmeester HJ, Sun Z, Alvarez-Huerta M, Verhoeven HA, Blaas J, van Houwelingen AMML, de Vos RCH, van der Voet H, Jansen RC, Guis M, Mol J, Davis RW, Schena M, van Tunen AJ and O'Connell AP. 2000. Identification of the *SAAT* gene involved in strawberry flavor biosynthesis by use of DNA microarrays. *Plant Cell* **12**, 647-661.

Åhman I. 1997. Growth, herbivory and disease in relation to gender in *Salix viminalis* L. *Oecologia* **111**, 61-68.

Åhman I and Larsson, S. 1994. Genetic improvement of willow (Salix) as a source of bioenergy. *Norwegian Journal of Agricultural Sciences*, supplement 18, 47-56.

AlonsoBlanco C, Peters AJM, Koornneef M, Lister C, Dean C, van den Bosch N, Pot J and Kuiper MTR. 1998. Development of an AFLP based linkage map of Ler, Col, and Cvi *Arabidopsis thaliana* ecotypes and construction of a Ler/Cvi recombinant inbred line population. *Plant Journal* **14**, 259-271.

Anderson JA, Sorrells ME and Tanksley SC. 1993. RFLP analysis of genomic regions associated with preharvest sprouting in wheat. *Crop Science* **33**, 453-459.

Anon. 1997. Communication from the Commission. Energy for the future: renewable sources of energy. White paper for a Community Strategy, COM (97) 299 final (Brussels 26.11.1997).

Anon. 2002. Geneticists branch out to sequence the poplar genome. *Nature* **415**, 725.

Arcade A, Anselin F, Faivre Rampant P, Lesage MC, Pâques LE and Prat D. 2000. Application of AFLP, RAPD and ISSR markers to genetic mapping of European and Japanese larch. *Theoretical and Applied Genetics* **100**, 299-307.

Argus GW. 1974. An experimental study of hybridisation and pollination in *Salix* (willow). *Canadian Journal of Botany* **52**, 1613-1619.

Arondel V, Lemieux B, Hwang I, Gibson S, Goodman HM and Somerville CR. 1992. Map-based cloning of a gene controlling omega-3-fatty acid desaturation in *Arabidopsis*. *Science* **258**, 1353-1355.

Asíns MJ. 2002. Present and future of quantitative trait locus analysis in plant breeding. *Plant Breeding* **121**, 281-299.

Atienza SG, Satovic Z, Peterson KK, Dolstra O and Martín A. 2003. Identification of QTLs influencing agronomic traits in *Miscanthus sinensis* Anderss. I. Total height, flag-leaf height and stem diameter. *Theoretical and Applied Genetics* **107**, 123-129.

Bachmann K. 1994. Molecular markers in plant ecology. *New Phytologist* **126**, 403-418.

Barker JHA, Matthes M, Arnold GM, Edwards KJ, Åhman I, Larsson S and Karp A. 1999. Characterisation of genetic diversity in potential biomass willows (*Salix* spp.) by RAPD and AFLP analyses. *Genome* **42**, 173-183.

Barker JHA, Pahlich A, Trybush S, Edwards KJ, Karp A. 2003. Microsatellite markers for diverse *Salix* species. *Molecular Ecology Notes* **3**, 4-6.

Barnard GA. 1949. Statistical inference. *Journal of the Royal Statistical Society Series B* **39**, 1-38.

Barreneche T, Bodenes C, Lexer C, Trontin JF, Fluch S, Streiff R, Plomiorobin C, Roussel G, Steinkellner H, Burg K, Favre JM, Glossl J and Kremer A. 1998. A genetic linkage map of *Quercus robur* L. (pedunculate oak) based on RAPD, SCAR, microsatellite, minisatellite, isozyme and 5S rDNA markers. *Theoretical and Applied Genetics* **97**, 1090-1103.

Barros E, Verryn S and Hettasch M. 2002. Identification of PCR-based markers linked to wood splitting in *Eucalyptus grandis*. *Annals of Forst Research* **59**, 675-678.

Bassoi M. 2002. Quantitative trait analysis of grain dormancy in wheat. PhD thesis, University of East Anglia, UK.

Basten C, Weir B, and Zhao-Bang Z. *QTL Cartographer*. (http://statgen.ncsu.edu/qtlcart/cartographer.html).

Beavis WD. 1994. The power and deceit of QTL experiments: lessons from comparative QTL studies. *Proceedings of the 49th Annual Corn and Sorghum Industry Research Conference*, Washington DC, ASTA.

Bent AF. 1996. Plant disease resistance genes: function meets structure. *Plant Cell* **8**, 1757-1771.

Bernacchi D, Beck-Bunn T, Eshed Y, Lopez J, Petiard V, Uhlig J, Zamir D and Tanksley S. 1998. Advanced backcross QTL analysis in tomato. I. Identification of QTLs for traits of agronomic importance from *Lycopersicon hirsutum*. *Theoretical and Applied Genetics* **97**, 381-397.

Bonierbale MW, Plaisted RL and Tanksley SD. 1988. RFLP maps based on common setes of clones reveal modes of chromosome evolution in potato and tomato. *Genetics* **120**, 1095-1103.

Botstein D, White, RL, Skolnick M and Davis RW. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* **32**, 314-331.

Bradshaw Jr HD, Villar M, Watson BD, Otto KG, Stewart S and Stettler RF. 1994. Molecular genetics of growth and development in *Populus.* III. A genetic linkage map of a hybrid poplar composed of RFLP, STS and RAPD markers. *Theoretical and Applied Genetics* **89**, 167-178.

Bradshaw Jr HD and Stettler RF. 1994. Molecular genetics of growth and development in *Populus.* II. Segregation distortion due to genetic load. *Theoretical and Applied Genetics* **139**, 963-973.

Bradshaw HD and Stettler R. 1995. Molecular genetics of growth and development in Populus. IV. Mapping QTLs with large effects on growth, form and phenology traits in a forest tree. *Genetics* **139**, 963-973.

Breese EL and Mather K. 1957. The organisation of polygenic variation within a chromosome in *Drosophila.* 1. Hair characters. *Heredity* **11**, 373-395.

Brenig B and Brem G. 1991. DNA cloning of sequence-tagged microsatellites sites by DNA affinity chromatography. *Nucleic Acids Research* **19**, 5441.

Bretting PK and Widrlechner MP. 1995. Genetic markers and horticultural germplasm management. *Hortscience* **30**, 1349-1356.

Britt CP, Gavaland A, Kofman PD, Ledin S, Sinner H-U and Spinelli R. 1997. The agronomy of short-rotation coppice - current and recent European research. *Aspects of Applied Biology* **49**, 33-40.

Brugmans B, van der Hulst RGM, Visser RGF, Lindhout P and van Eck HJ. 2003. A new and versatile method for the successful conversion of AFLP markers into simple single locus markers. *Nucleic Acids Research* **31**, article number e55.

Büchler W. 1986. Neue chromosomomenzählungen in der gattung *Salix. Botanica Helvetica* **95**, 165-175.

Büchler W. 1986. Neue chromosomomenzählungen in der gattung *Salix.* 2. Teil. *Botanica Helvetica* **96**, 135-143.

Burr B and Burr FA. 1991. Recombinant inbreds for molecular mapping in maize: theoretical and practical considerations. *Trends in Genetics* **7**, 55-60.

Byrne PF and McMullen MD. 1996. Defining genes for agricultural traits: QTL analysis and the candidate gene approach. *Probe* **7**, 24-27.

Caetano-Anollés G, Schlarbaum SE and Trigiano RN. 1999. DNA amplification fingerprinting and marker screening for pseudo-testcross mapping of flowering dogwood (*Cornus florida* L.). *Euphytica* **106**, 209-222.

Cai Q, Guy CL and Moore GA. 1994. Extension of the linkage map in Citrus using random amplified polymorphic DNA (RAPD) markers and RFLP mapping of cold-acclimation-responsive loci. *Theoretical and Applied Genetics* **89**, 606-614.

Cato SA, Gardner RC, Kent J and Richardson TE. 2001. A rapid PCR-based method for genetically mapping ESTs. *Theoretical and Applied Genetics* **102**, 296-306.

Causse M, Saliba V, Philouze J, Buret M, Chevalier C and Rothan C. 1999. Mapping QTLs and genes involved in tomato fruit weight and sugar content. In: *Proceedings of the Molecular Biology of the Tomato Congress*, York, UK.

Cervera MT, Storme V, Ivens B, Gusmão J, Liu BH, Hostyn V, Van Slycken M, Van Montague M and Boerjan W. 2001. Dense genetic linkage maps of three *Populus* species (*Populus deltoides*, *P. nigra* and *P. trichocarpa*) based on AFLP and microsatellite markers. *Genetics* **158**, 787-809.

Cevik V and King GJ. 2002. High-resolution genetic analysis of the Sd-1 aphid resistance locus in *Malus* spp. *Theoretical and Applied Genetics* **105**, 346-354.

Chagne D, Lalanne C, Madur D, Kumar S, Frigerio JM, Krier C, Decroocq S, Savoure A, Bou-Dagher-Kharrat M, Bertocchi E, Brach J and Plomion C. 2002. A high density genetic map of maritime pine based on AFLPs. *Annals of Forest Research* **59**, 627-636.

Chang C, Kwok SF, Bleecker AB, Meyerowitz EM. 1993. *Arabidopsis* ethylene-response gene *ETR1* – similarity of product to 2-component regulators. *Science* **262**, 39-544.

Chao S, Sharp PJ, Gale MS. 1988. A linkage map of wheat homoeologous group 7 using RFLP markers. In: *Proceedings of the 7th International Wheat Genetics Symposium.* (eds. Miller TE and Koebner RMD), IPSR, Cambridge Laboratory, Cambridge, 493-498.

Cho RJ, Mindrinos M, Richards DF, Sapolsky RJ, Anderson M, Drenkard E, Dewdney J, Reuber TL, Stammers M, Federspiel N, Theologis A, Yang W, Hubbell E, Au M, Chung EY, Lashkari D, Lemieux B, Dean C, Lipshutz RJ, Ausubel FM , Davis RW and Oefner PJ. 1999. Genome-wide mapping with biallelic markers in *Arabidopsis thaliana. Nature Genetics* **23**, 203-207.

Choo TM. 1981. Doubled haploids for studying the inheritance of quantitative characters. *Genetics* **99**, 525-540.

Churchill GA and Doerge RW. 1994. Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963-971.

Collins CM. 1999. Salicine vampires? Aspects of the biology and ecology of the giant willow aphid, *Tuberolachnus salignus*, and the black willow aphid, *Pterocomma salicis. Antenna* **23**, 99-100.

Coltman DW, Bowen WD and Wright JM. 1996. PCR primers for harbour seal (*Phoca vitulina concolour*) microsatellites amplify polymorphic loci in other pinniped species. *Molecular Ecology* **5**, 161-163.

Corbett SJ and Britt CP. 1997. Invertebrate pest monitoring on arable energy coppice: a one year study in south Warwickshire. *Aspects of Applied Biology* **49**, 149-154.

Cregan PB, Mudge J, Fickus EW, Danesh D, Denny R and Young RD. 1999. Two simple sequence repeat markers to select for soybean cyst nematode resistance conditioned by the rhg1 locus. *Theoretical and Applied Genetics* **99**, 811-818.

Cullis BR and Gleeson AC. 1991. Spatial analysis of field experiments - and extension to two dimensions. *Biometrics* **47**, 1449-1460.

Cullis BR, Gogel BJ, Verbyla A and Thompson R. 1998. Spatial analysis of multi-environment early generation variety trials. *Biometrics* **54**, 1-19.

Darvasi A, Weintreb A, Minke V, Weller J, Soller M. 1993. Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics* **134**, 943-951.

Dawson WM. 1992. Some aspects of the development of short-rotation coppice willow for biomass in Northern Ireland. *Proceedings of the Royal Society of Edinburgh* **98B**, 193-205.

Dawson WM and McCracken AR. 1994. Effect of *Melampsora* rust on growth and development of *Salix burjatica* 'Korso' in Northern Ireland. *European Journal of Forest Pathology* **24**, 32-39.

Debener T and Mattiesch L. 1999. Construction of a genetic linkage map for roses using RAPD and AFLP markers. *Theoretical and Applied Genetics* **99**, 891-899.

Denno RF, Larsson S and Olmstead KL. 1990. Role of enemy-free space and plant quality in host-plant selection by willow beetles. *Ecology* **71**, 124-137.

Devos KM, Atkinson MD, Chinoy CN, Harcourt RL, Koebner RMD, Liu CJ, Masojc P, Xie DX and Gale MD. 1993. Chromosomal rearrangements in the rye genome relative to that of wheat. *Theoretical and Applied Genetics* **85**, 673-680.

Devos KM and Gale MD. 1997. Comparative genetics in the grasses. *Plant Molecular Biology* **35**, 3-15.

Devos KM and Gale MD. 2000. Genome relationships: the grass model in current research. *Plant Cell* **12**, 637-646.

Doerge RW and Chrurchill GA. 1996. Permutation tests for multiple loci affecting a quantitative character. *Genetics* **142**, 285-294.

Doerge RW and Rebaï A. 1996. Significance thresholds for QTL interval mapping tests. *Heredity* **77**, 459-464.

Doligez A, Bouquet A, Danglot Y, Lahogue F, Riaz S, Meredith CP, Edwards KJ and This P. 2002. Genetic mapping of grapevine (*Vitis vinifera* L.) applied to the detection of QTLs for seedlessness and berry weight. *Theoretical and Applied Genetics* **105**, 780-795.

Donini P, Stephenson P, Bryan GJ and Koebner RMD. 1998. The potential of microsatellites for high-throughput genetic diversity assessment in wheat and barley. *Genetic Resources and Crop Evolution* **45**, 415-421.

Dumbleton F. 1997. Biomass conversion technologies – an overview. *Aspects of Applied Biology* **49**, 341-347.

Dupuis J and Siegmund D. 1999. Statistical methods for mapping quantitative trait loci from a dense set of markers. *Genetics* **151**, 373-386.

Duvick DN. 1992. Genetic contributions to advances in yield of United States maize. *Maydica* **37**, 69-79.

Edwards A, Civitello A, Hammond HA and Caskey CT. 1991. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *American Journal of Human Genetics* **49**, 746-756.

Edwards KJ, Barker JHA, Daly A, Jones C and Karp A. 1996. Microsatellite libraries enriched for several microsatellite sequences in plants. *Biotechniques* **20**, 758-760.

Emerson RA, Beadle GW and Fraser AC. 1935. A summary of linkage studies in maize. *Cornell University Agricultural Experimental Station Memoir* **180**.

Etienne C, Rothan C, Moing A, Svanella L, Dirlewanger E, Just D and Monet R. 1999. Isolation of peach (*Prunus persica* L. Batsch) cDNAs expressed during fruit development. In Heller SR (ed.), International Conference on the Status of Plant and Animal Genome Research VII. Abstracts. San Diego, CA, USA.

Falconer DS. 1960. *Introduction to Quantitative Genetics*, 1st Edn. Oliver and Boyd, Edinburgh.

Feingold E, Brown PO and Siegmund D. 1993. Gaussian models for genetic linkage analysis using complete high-resolution maps of identity of descent. *American Journal of Human Genetics* **53**, 234-251.

Fischer SG and Lerman LS. 1983. DNA fragments differing by single base-pair substitutions are separated in denaturing gradient gels: correspondence with melting theory. *Proceedings of the National Academy of Science, USA* **80**, 1579-1583.

Fisher RA. 1918. The correlations between relative supposition of Mendelian inheritance. *Transcripts of the Royal Society of Edinburgh* **52**, 399-433.

Flintham JE. 2000. Different genetic components control coat-imposed and embryo-imposed dormancy in wheat. *Seed Science Research* **10**, 43-50.

Fourmann M, Barret P, Renard M, Pelletier G, Delourme R and Brunel D. 1998. The two genes homologous to *Arabidopsis FAE1* co-segregate with the two loci governing erucic acid content in *Brassica napus*. *Theoretical and Applied Genetics* **96**, 852-858.

Frewen BE, Chen THH, Howe GT, Davis J, Rohde A, Boerjan W and Bradshaw HD. 2000. Quantitative trait loci and candidate gene mapping of bud set and bud flush in *Populus*. *Genetics* **154**, 837-845.

Frary A, Nesbitt TC, Grandillo S, Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB and Tanksley SD. 2000. *fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. *Science* **289**, 85-88.

Freeman WM, Walker SJ and Vrana KE. 1999. Quantitative RT-PCR: Pitfalls and potential. *Biotechniques* **26**, 112.

Fridman E, Pleban T and Zamir D. 2000. A recombinant hotspot delimits a wild QTL for tomato sugar content to 484 bp within an invertase gene. *Proceedings of the National Academy of Sciences, USA* **97**, 4718-4723.

Gale MD and Devos KM. 1998a. Plant comparative genetics after 10 years. *Science* **282**, 656-659.

Gale MD and Devos KM. 1998b. Comparative genetics in the grasses. *Proceedings of the National Academy of Sciences, USA* **95**, 1971-1974.

Gale MD, Flintham JE and Devos KM. 2002. Cereal comparative genetics and preharvest sprouting. *Euphytica* **126**, 21-25.

Galton F. 1889. *Natural Inheritance.* Macmillan, London.

Gebhardt C, Ritter E, Barone A, Debener T, Walkemeier B, Schachtschabel U, Kaufmann H, Thompson RD, Bonierbale MW, Ganal MW, Tanksley SD and Salamini F. RFLP maps of potato and their alignment with the homoeologous tomato genome. *Theoretical and Applied Genetics* **83**, 49-57.

Gelderman H. 1975. Investigation on inheritance of quantitative characters in animals by gene markers. I. Methods. *Theoretical and Applied Genetics* **46**, 300-319.

Gemmell NJ, Allen PJ, Goodman SJ and Reed JZ. 1997. Interspecific microsatellite markers for the study of pinniped populations. *Molecular Ecology* **6**, 661-666.

Gentzbittel L, Mouzeyar S, Badaoui S, Mestries E, Vear F, Tourvielle de Labrouhe D and Nicolas P. 1998. Cloning of markers for disease resistance in sunflower, *Helianthus annuus* L. *Theoretical and Applied Genetics* **96**, 519-525.

Genstat Committee (1993) Genstat 5 Release 3 Reference Manual, Clarendon Press, Oxford, UK.

Giovanonni J. 2000. Tools for tomato genome analysis: application for isolation of fruit ripening loci and expression profiling. In: Heller SR (ed.) International Conference on the Status of Plant and Animal Research VIII. Abstracts. San Diego, CA, USA.

Gleeson AC and Cullis BR. 1987. Residual Maximum Likelihood (REML) estimation of a neighbour model for field experiments. *Journal of Agricultural, Biological and Environmental Statistics* **2**, 1-25.

Glynn C. 1996. Costs in the interactions between a gall midge and willow. Swedish University of Agricultural Sciences, Uppsala, Sweden.

Gosselin I, Zhou Y, Bousquet J and Isabel N. 2002. Megagametophyte-derived linkage maps of white spruce (*Picea glauca*) based on RAPD, SCAR and ESTP markers. *Theoretical and Applied Genetics* **104**, 987-997.

Grattapaglia D and Sederoff R. 1994. Genetic linkage maps of *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross strategy and RAPD markers. *Genetics* **137**, 1121-1137.

Green JM, Batley J, Peacock L, Carter P, Glen DM and Karp A. 2001. The molecular ecology of willow beetles (*Phyllodecta* spp.). *Antennae* **25**, 254-257.

Gupta PK, Balyan HS, Sharma PC and Ramesh R. 1996. Microsatellites in plants: A new class of molecular markers. *Current Science* **70**, 45-54.

Gut I. 2001. Automation in genotyping of single nucleotide polymorphisms. *Human Mutation* **17**, 475-492.

Håkansson A. 1955. Chromosome numbers and meiosis in certain salices. *Hereditas* **41**, 454-482.

Haldane JBS. 1919. The combination of linkage values, and the calculation of distance between the loci of linked factors. *Journal of Genetics* **8**, 299-309.

Haley CS and Andersson L. 1997. Linkage mapping of quantitative trait loci in plants and animals. In: *Genome mapping – a practical approach*, 49-71. (ed: Dear PH). IRL Press at Oxford University Press.

Haley CS and Knott SA. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**, 315-324.

Hall DO, Scrase JI and Rosillo-Calle F. 1997. Biomass energy: the global context now and in the future. *Aspects of Applied Biology* **49**, 1-10.

Hämäläinen JH, Sorri VA, Watanabe KN, Gebhardt C and Valkonen JPT. 1998. Molecular examination of a chromosome region that controls resistance to potato Y and A potyviruses in potato. *Theoretical and Applied Genetics* **96**, 1036-1043.

Han F, Romagosa I, Ullrich SE, Jones BL, Hayes PM and Wesenberg DM. 1997. Molecular marker-assisted selection for malting quality traits in barley. *Molecular Breeding* **3**, 247-437.

Hanley S, Edwards D, Stevenson D, Haines S, Hegarty M, Schuch W and Edwards KJ. 2000. Identification of transposon-tagged genes by the random sequencing of *Mutator*-tagged DNA fragments from *Zea mays*. *The Plant Journal* **22**, 557-566.

Hanley S, Barker JHA, Van Ooijen JW, Aldam C, Harris SL, Ahman I, Larsson S and Karp A. 2002. A genetic linkage map of willow (*Salix viminalis*) based on AFLP and microsatellite markers. *Theoretical and Applied Genetics* **105**, 1087-1096.

Harry DE, Temesgen B and Neale DB. 1998. Co-dominant PCR-based markers for *Pinus taeda* developed from mapped cDNA clones. *Theoretical and Applied Genetics* **98**, 327-336.

Hemmat M, Weedon NF, Manganaris AG and Lawson DM. 1994. A molecular marker linkage map for apple. *Journal of Heredity* **85**, 4-11.

Hilbert P, Lindpainter K, Beckmann JS, Seriwaka T, Soubrier F, Dubay C, Cartwright P, de Gouyon B, Julier C, Takahasi S, Vincent M, Ganten D, Georges M and Lanthrop GM. 1991. Chromosomal mapping of two genetic loci associated with blood-pressure regulation in hereditary hypersensitive rats. *Nature* **353**, 521-529.

Hunter T and Peacock L. 2001. Effects of host genotype on insect and disease damage and yields in SRC willows. *Aspects of Applied Biology* **65**, 263-270.

Hutchinson HP and Kearns HGH. 1930a. The control of *Galerucella lineola* - a major pest of willows. *Annual Report of Agricultural and Horticultural Research Station, Long Ashton* pp108-111.

Hutchinson HP and Kearns HGH. 1930b. The control of *Phyllodecta vitellinae* (L.) a major pest of willows. *Annual Report of Agricultural and Horticultural Research Station, Long Ashton* pp112-126.

Hyne V and Kearsey MJ. 1995. QTL analysis: Further uses of marker regression. *Theoretical and Applied Genetics* **91**, 471-476.

Hyne V, Kearsey MJ, Pike DJ and Snape JW. 1995. QTL analysis: unreliability and bias in estimation procedures. *Molecular Breeding* **1**, 273-282.

Iguartua E, Edney M, Rossnagel BG, Spaner D, Legge WG, Scoles GJ, Eckstein PE, Penner GA, Tinker NA, Briggs KG, Falk DE and Mather DE. 2000. Marker-based selection of QTL affecting grain and malt quality in two-row barley. *Crop Science* **40**, 1426-1433.

Jacob HJ, Lindpaintner K, Lincoln SE, Kusumi K, Bunker, RK, Mao Yi-Pei, Ganten D, Dzau VJ and Lander, ES. 1991. Genetic mapping of a gene causing hypertensive rat. *Cell* **67**, 213-224.

Jarne P and Lagoda PJL. 1996. Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution* **11**, 424-429.

Jansen RC. 1993. Interval mapping of multiple quantitative trait loci. *Genetics* **135**, 205-211.

Jansen RC. 1994. Controlling the type I and type II errors in mapping quantitative trait loci. *Genetics* **138**, 871-881.

Jansen RC and Stam P. 1994. High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* **136**, 1447-1455.

Jansen RC, Van Ooijen JW, Stam P, Lister C and Dean C. 1995. Genotype-by-environment interaction in genetic mapping of multiple trait loci. *Theoretical and Applied Genetics* **91**, 33-37.

Jeffreys AJ, Allen MJ, Armour JAL, Collick A, Dubrova Y, Fretwell N, Guram T, Jobling M, May CA, Neil DL and Neumann R. 1995. Mutation processes at human minisatellites. *Electrophoresis* **16**, 1577-1585.

Jeffreys AJ, Royle NJ, Wilson V and Wong Z. 1988. Spontaneous mutation rates to new length alleles at tandem repetitive hypervariable loci in human DNA. *Nature* **332**, 278-281.

Jones CJ, Edwards KJ, Castaglione S, Winfield MO, Sala F, Van de Wiel C, Bredemeijer G, Vosman B, Matthes M, Daly A, Brettschneider R, Bettini P, Buiatti M, Maestri E, Malcevschi A, Marmiroli N, Aert R, Volckaert G, Rueda J, Linacero R, Vazquez A and Karp A. 1997. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Molecular Breeding* **3**, 381-390.

Jones N, Ougham H and Thomas H. 1997. Markers and mapping: we are all geneticists now. *New Phytologist* **137**, 165-177.

Julier C, Delepine M, Keavney B, Terwilliger J, Davis S, Weeks DE, Bui T, Jeunemaitre X, Velho G, Froguel P, Ratcliffe P, Corvol P, Soubrier F and Lathrop GM. 1997. Genetic susceptibility for human familial essential hypertension in a region of homology with blood pressure linkage on rat chromosome 10. *Human Molecular Genetics*, **6**, 2077-2085.

Kanazin V, Marek L and Shoemaker RC. Resistance gene analogues are conserved and clustered in soybean. *Proceedings of the National Academy of Science, USA* **93**, 11746-11750.

Karagyozov L, Kolcheva ID and Chapman VM. 1993. Construction of random small-insert genomic libraries highly enriched for simple sequence repeats. *Nucleic Acids Research* **21**, 3911-3912.

Karp A, Seberg O and Buiatti M. 1996. Molecular techniques in the assessment of botanical diversity. *Annals of Botany* **78**, 143-149.

Karp A, Edwards KJ, Bruford M, Funk S, Vosman B, Morgante M, Seberg O, Kremer A, Boursot P, Arctander P, Tautz D and Hewitt GM. 1997. Molecular technologies for biodiversity evaluation: opportunities and challenges. *Nature Biotechnology* **13**, 625-628.

Karp A and Peacock L. 2003. The ecology and population genetics of the blue and brassy willow beetles (*Phratora (=Phyllodecta) vulgatissima* L. and P. *vitellinae* L.) on UK willow plantations. In: *New contributions to the biology of Chrysomelidae.* (eds. Jolivet PH, Santiago-Blay JA and Schmitt, M.) (In press).

Kashi Y, King D and Soller M. 1997. Simple sequence repeats as a source of quantitative genetic variation. *Trends in Genetics* **13**, 74-78.

Kearsey M. 1998. The principles of QTL analysis (a minimal mathematics approach). *Journal of Experimental Botany* **49**, 1619-1623.

Kearsey MJ and Farquhar GL. 1998. QTL analysis in plants; where are we now? *Heredity* **80**, 137-142.

Kearsey MJ and Hyne V. 1994. QTL analysis: a simple 'marker regression' approach. *Theoretical and Applied Genetics* **89**, 698-702.

Kearsey M and Pooni. 1996. The genetical analysis of quantitative traits. Pub: Chapman and Hall, London.

Kelly MT and Curry JP. 1991. The biology and population density of the willow beetle (*Phratora vulgatissima* [L.]) on *Salix viminalis* in reclaimed cutaway peat. *Journal of Applied Entomology* **111**, 44-56.

Kendall DA, Hunter T, Arnold GM, Liggitt J, Morris T and Wiltshire CW. 1996. Susceptibility of willow clones (*Salix* spp) to herbivory by *Phyllodecta vulgatissima* (L) and *Galerucella lineola* (Fab4) (Coleoptera, Chrysomelidae). *Annals of Applied Biology* **129**, 379-390.

Kendall DA and Wiltshire CW. 1998. Life-cycles and ecology of willow beetles on *Salix viminalis* in England. *European Journal of Forest Pathology* **28**, 281-288.

Kennedy CEJ and Southwood TRE. 1984. The number of species of insects associated with British tress: a re-analysis. *Journal of Animal Ecology* **53**, 455-478.

Kermicle JL. 1969. Androgenesis conditioned by a mutation in maize. *Science* **166**, 1422-1424.

King GJ. 2002. Through a genome, darkly: comparative analysis of plant chromosomal DNA. *Plant Molecular Biology* **48**, 5-20

Knott SA and Haley CS. 1992. Maximum likelihood mapping of quantitative trait loci using full-sib families. *Genetics* **132**, 1211-1222.

Kolehmainen J, Julkunen-Tiitto R, Roininen H and Tahvanainen J. 1995. Phenolic glucosides as feeding cues for willow-feeding leaf beetles. *Entomologia Experimentalis et Applicata* **74**, 235-243.

Kondo T, Terada K, Hayashi E, Kuramoto N, Okamura M and Kawasaki H. 2000. RAPD markers linked to a gene for resistance to pine needle gall midge in Japanese black pine (*Pinus thunbergii*). *Theoretical and Applied Genetics* **100**, 391-395.

Koornneef M, Alonso-Blanco C, Peeters AJM and Soppe W. 1998. Genetic control of flowering time in *Arabidopsis*. *Annual Review of Plant Physiology and Plant Molecular Biology*, **49**, 345-370.

Kosambi DD. 1955. The estimation of map distances from recombination values. *Annals of Eugenics* **12**, 172-175.

Kowalski SP, Lan TH, Feldmann KA and Paterson AH. 1994. Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveal islands of conserved organization. *Genetics* **138**, 499-510.

Kresovich S and McFerson JR. 1992. Assessment and management of plant genetic diversity - considerations of intraspecific and interspecific variation. *Field Crops Research* **29**, 185-204.

Kruglyak L. 1997. The use of a genetic map of biallelic markers in linkage studies. *Nature Genetics* **17**, 21-24.

Kruglyak L and Lander ES. 1995. A nonparametric approach for mapping quantitative trait loci. *Genetics* **139**, 1421-1428.

Krutovskii KV, Vollmer SS, Sorensen FC, Adams WT, Knapp SJ and Strauss SH. 1998. RAPD genome maps of Douglas-fir. *Journal of Heredity* **89**, 197-205.

Kubisiak TL, Nelson CD, Nance WL and Stine M. 1995. RAPD linkage mapping in a longleaf pine x slash pine F_1 family. *Theoretical and Applied Genetics* **90**, 1119-1127.

Kumar LS, 1999. DNA markers in plant improvement: An overview. *Biotechnology Advances* **17**, 143-182.

Kwok PY. 2000. High-throughput genotyping assay approaches. *Pharmacogenomics* **1**, 95-100.

Kwok PY, Deng Q, Zakeri H, Taylor SL and Nickerson DA. 1996. Increasing the information content of STS-based genome maps: identifying polymorphisms in mapped STSs. *Genomics* **31**, 123-126.

la Rosa R, Angiolillo A, Guerrero C, Pellegrini M, Rallo L, Besnard G, Berville A, Martin A and Baldoni L. 2003. A first linkage map of olive (*Olea Europea* L.) cultivars using RAPD, AFLP, RFLP and SSR markers. *Theoretical and Applied Genetics* **106**, 1273-1282.

Lagercrantz U, Putterill J, Coupland G and Lydiate D. 1996. Comparative mapping in *Arabidopsis* and *Brassica*, fine scale genome collinearity and congruence of genes controlling flowering time. *Plant Journal* **9**, 13-20.

Lan TH and Paterson AH. 2000. Comparative mapping of QTL sculpting the curd of *Brassica oleracea. Genetics* **155**, 1927-1954.

Lan TH and Paterson AH. 2001. Comparative mapping of QTLs determining the plant size of *Brassica oleracea. Theoretical and Applied Genetics* **103**, 383-397.

Lanaud C, Risterucci AM, Ngoran AKJ, Clement D, Flament MH, Laurent V and Falque M. 1995. A genetic linkage map of *Theobroma cacao. Theoretical and Applied Genetics* **91**, 987-993.

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE and Newburg L. 1987. MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**, 174-181.

Lander ES and Botstein D. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185-199.

Lander ES and Kruglyak L. 1995. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genetics* **11**, 241-247.

Larsson S. 1996. Willow coppice as short rotation forestry. In: *Energy from Crops.* Semundo Limited, Cambridge, UK.

Larsson S. 2001. Commercial varieties from the Swedish willow breeding programme. *Aspects of Applied Biology* **65**, 193-198.

Lefèvre FC, Pichot C and Pinon J. 1994. Intra- and interspecific inheritance of some components of the resistance to leaf rust (*Melampsora larici-populina* Kleb.) in poplars. *Theoretical and Applied Genetics* **88**, 501-507.

Lefebvre V, Kuntz M, Camara B and Plloix A. 1998. The capsanthin-capsorubin synthase gene: a candidate gene for the *y* locus controlling red fruit colour in pepper. *Plant Molecular Biology* **36**, 785-789.

Lee M. 1995. DNA markers and plant breeding programs. *Advances in Agronomy* **55**, 265-344.

Lee M. 1998. Genome projects and gene pools: New germplasm for plant breeding? *Proceedings of the National Academy of Sciences USA* **95**, 2001-2004.

Lehmann EL. 1975. *Nonparametrics.* McGraw-Hill, New York.

Leister D, Kurth J, Laurie DA, Yano M, Sasaki T, Devos K, Graner A and Schulze-Lefert P. 1998. Rapid reorganization of resistance gene homologues in cereal genomes. *Proceedings of the National Academy of Sciences, USA* **95**, 370-375.

Lespinasse D, Rodier-Gous M, Grivet L, Leconte A, Legnate H and Seguin M. 2000. A saturated genetic linkage map of rubber tree (*Hevea* spp.) based on RFLP, AFLP, microsatellite and isozyme markers. *Theoretical and Applied Genetics* **100**, 127-138.

Leung J, Bouvierdurand M, Morris PC, Guerrier D, Chefdor F and Giraudat J. 1994. Arabidopsis ABA response gene *ABI1* – features of calcium-modulated protein phosphatase. *Science* **264**, 1448-1452.

Lin D, Hubbes M and Zsuffa L. 1994. Differentiatition of willow and poplar clones using RAPD fingerprints. *Tree Physiology* **14**, 1097-1105.

Lincoln, S, Daly M, and Lander E. 1992. Mapping genes controlling quantitative traits with MAPMAKER/QTL 1.1. Whitehead Institute Technical Report, 2nd edition.

Lindegaard KN. 2002. Breeding strategies – the way forward. DTI seminar series, Blagdon, Somerset. In press.

Lindegaard KN and Barker JHA. 1997. Breeding willows for biomass. *Aspects of Applied Biology*, **49**, 155-162.

Lindegaard KN, Parfitt RI, Donaldson G, Hunter T, Dawson WM, Forbes EGA, Carter MM, Whinney CC, Shinney JE and Larsson S. 2001. Comparative trials of elite Swedish and UK biomass varieties. *Aspects of Applied Biology* **65**, 183-192.

Liu BH. 1997. *Statistical genomics.* CRC press, Boca Raton, New York.

Livingstone KD, Lackney VK, Blauth JR, van Wijk R and Jahn MK. 1999. Genome mapping in *Capsicum* and the evolution of genome structure in the *Solanaceae. Genetics* **152**, 1183-1202.

Litt M and Luty JA. 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac-muscle actin gene. *American Journal of Human Genetics* **44**, 397-401.

Lynch M and Walsh B. 1998. *Genetics and analysis of quantitative traits.* Sinauer Associates, Sunderland, Massachussetts.

MacArthur JW. 1934. Linkage groups in tomato. *Journal of Genetics* **29**, 123-133.

Maliepaard C and Van Ooijen JW. 1994. QTL mapping in a full-sib family of an outcrossing species. In: Van Ooijen JW , Jansen J (eds) Biometrics in plant breeding: applications of molecular markers. *Proceedings of the 9th Meeting EUCARPIA Section Biometrics and Plant Breeding.* Plant Reseearch International, Wageningen, The Netherlands, 140-146.

Maliepaard C, Jansen J and Van Ooijen JW. 1997. Linkage analysis in a full-sib family of an outbreeding plant species: overview and consequences for applications. *Genetical Research, Cambridge* **70**, 237-250.

Maliepaard C, Alston FH, van Arkel G, Brown LM, Chevreau E, Dunemann F, Evans KM, Gardiner S, Guilford P, van Heusden AW, Janse J, Laurens F, Lynn JR, Manganaris AG, den Nijs APM, Periam N, Rikkerink E, Roche P, Ryder C, Sansavini S, Schmidt H, Tartarini S, Verhaegh JJ, Vrielink-van Ginkel M and King G. 1998. Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. *Theoretical and Applied Genetics* **97**, 60-73.

Manly KF and Olson JM. 1999. Overview of QTL mapping software and introduction to map manager QTs. *Mammalian Genome* **10**, 327-334.

Markert CL and Moller F. 1959. Multiple forms of enzyme: tissue, ontogenic and species specific patterns. *Proceedings of the National Academy of Sciences USA,* **45**, 753–763.

Marques CM, Araújo JA, Ferreira JG, Whetten R, O'Malley DM, Liu B-H and Sederoff R. 1998. AFLP genetic maps of *Eucalyptus globulus* and *E. tereticornis. Theoretical and Applied Genetics* **96**, 727-737.

Marques CM, Vasquez-Kool J, Carocha VJ, Ferreira JG, O'Malley DM, Liu BH and Sederoff R. 1999. Genetic dissection of vegetative propagation traits in *Eucalyptus tereticornis* and *E. globules. Theoretical and Applied Genetics* **99**, 936-946.

Martin GB, Brommonschenkel SH, Chunwongse J, Frary A, Ganal MW, Spivey R, Wu TY, Earle ED and Tanksley SD. 1993. Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science* **262**, 432-1436.

Mather K. 1949. *Biometrical Genetics*, 1st Edn. Methuen, London.

Mathes M, Daly A and Edwards KJ. 1998. Amplified fragment length polymorphisms. In: *Molecular tools for screening biodiversity: Plants and Animals.* (eds. Karp A, Isaac PG and Ingram DS), Chapman and Hall, London.

Mattson WJ. 1980. Herbivory in relation to plant nitrogen content. *Annual Review of Ecology and Systematics* **11**, 119-161.

Mazeyrat F, Mouzeyar S, Nicolas P, Tourvielle de Labrouhe D and Ledoight G. 1998. Cloning, sequence and characterization of a sunflower (*Helianthus annuus* L.) pathogen-induced gene showing sequence homology with auxin-induced genes in plants. *Plant Molecular Biology* **38**, 899-903.

Mazur BJ and Tingey SV. 1995. Genetic mapping and introgression of genes of agronomic importance. *Current Opinion in Biotechnology* **6**, 175-182.

McCouch SR, Kochert G, Yu ZH, Wang ZY, Khush GS, Coffman WR and Tanksley SD. 1988. Molecular mapping of rice chromosomes. *Theoretical and Applied Genetics* **76**, 815-829.

McCracken AR. 2001. Disease effects in mixed varietal plantations of willow. *Aspects of Applied Biology* **65**, 255-262.

McCracken AR and Dawson WM. 1997. Using mixtures of willow clones as a means of controlling rust disease. *Aspects of Applied Biology* **49**, 97-104.

Mead R, Curnow RN and Hasted, AM. 1993. *Statistical methods in agriculture and experimental biology*, 2nd edition, Chapman and Hall, London.

Melake-Berhan A, Hulbert SH, Butler LG, Bennetzen JL. 1993. Structure and evolution of the genomes of *Sorghum bicolor* and *Zea mays*. *Theoretical and Applied Genetics* **86**, 598-604.

Melchinger AH. 1990. Use of molecular markers in breeding for oligogenic disease resistance. *Plant Breeding - Zeitschrift fuer Pflanzenzuechtung* **104**, 1-19.

Melchinger AH, Boppenmaier J, Dhillon BS, Pollmer WG and Herrmann R. 1992. Genetic diversity for RFLPs in European maize inbreds: II: Relation to performance of hybrids within versus between heterotic groups for forage traits. *Theoretical and Applied Genetics* **84**, 672-681.

Melchinger AE, Utz HF and Schön CC. 1998. Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power QTL detection and large bias in estimates of QTL effects. *Genetics* **149**, 383-403.

Meyer K, Leube MP, Grill EA. 1994. Protein phosphatase 2C involved in ABA signal-transduction in *Arabidopsis thaliana*. *Science* **264**, 1452-1455.

Michelmore RW, Paran I and Kesseli RV. 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. *Proceeding of the National Academy of Sciences, USA* **88**, 9828-9832.

Mindrinos M, Katagiri F, Yu GL and Ausubel FM. 1994. The *A. thaliana* disease resistance gene RPS2 encodes a protein containing a nucleotide-binding site and leucine –rich repeats. *Cell* **78**, 1089-1099.

Mitchell SE, Kresovich S, Jester CA, Hernandez CJ and SzewcMcFadden AK. 1997. Application of multiplex PCR and fluorescence-based, semi-automated allele sizing technology for genotyping plant genetic resources. *Crop Science* **37**, 617-624.

Moncrief ND, Cockett NE, Neff AD, Thomas WL and Dueser RD. 1997. Polymorphic microsatellites in the meadow vole *Microtus pennysylvanicus* conservation of loci across species of rodents. *Molecular Ecology* **6**, 299-301.

Moreau L, Monod H, Charcosset and Gallais A. 1999. Marker-assisted selection with spatial analysis of unreplicated field trials. *Theoretical and Applied Genetics* **98**, 234-242.

Moretzsohn MC, Nunes CDM, Ferreira ME and Grattapaglia D. 2000. RAPD linkage mapping of the shell thickness locus in oil palm (*Elaeis guineensis* Jacq.). *Theoretical and Applied Genetics* **100**, 63-70.

Mudge J, Cregan PB, Kenworthy JP, Kenworthy WJ, Orf JH and Young ND. 1997. Two microsatellite markers that flank the major soybean cyst nematode resistance locus. *Crop Science* **37**, 1611-1615.

Newsholme C. 1992. Willows : The genus *Salix*. B.T. Batsford Ltd., London.

Nilsson NH. 1954. Uber hochkomplexe bastardverbindungen in der gattung *Salix*. *Heriditas* **40**, 517-522.

Nishikawa N, Oishi M and Kiyama R. 1995. Construction of a human genomic library of clones containing poly(dGdA):poly(dTdC) tracts by Mg^{2+}-dependent triplex affinity capture – DNA polymorphism associated with the tracts. *Journal of Biological Chemisrty* **270**, 9258-9264.

Nitzsche W and Wenzel G. 1977. *Haploids in plant breeding*. Paul Parey, Hamburg.

Nordin A and Kjellström B. 1996. Biomass for heat and electricity production – engineering aspects. In: *Energy from Crops*. Semundo Limited, Cambridge, UK.

Orita M, Iwahana H, Kanazawa H, Hayashi K and Sekiya T. 1989. Detection of polymorphisms of human DNA by gel electrophoresis as single-stranded conformational polymorphisms. *Proceedings of the National Academy of Sciences, USA* **86**, 2766-2770.

Ostrander EA, Jong PM, Rine J and Duyk G. 1992. Construction of small-insert genomic DNA libraries highly enriched for microsatellite repeat sequences. *Proceedings of the National Academy of Sciences USA* **89**, 3419-3423.

Palo RT. 1984. Distribution of birch (*Betula* spp.), willow (*Salix* spp.) and poplar (*Populus* spp.) secondary metabolites and their potential role as chemical defense against herbivores. *Journal of Chemical Ecology* **10**, 499-520.

Paran I and Michelmore RW. 1993. Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. *Theoretical and Applied Genetics*, **85**, 985-993.

Pasteels JM and Rowell-Rahier M. 1992. The chemical ecology of herbivory on willows. *Proceedings of the Royal Society of Edinburgh* **98B**, 63-73.

Paterson AH, Tanksley SD and Sorrells ME. 1991. DNA markers in plant improvement. *Advances in Agronomy* **46**, 39-90.

Paterson AH. 1996. Genome mapping in plants. R.G. Landers Company, Austin, Texas, USA.

Paterson AH. 1998. Of blending, beans, and bristles: The foundations of QTL mapping. In: *Molecular Dissection of Complex Traits*. (ed. Paterson AH) CRC Press, Boca Raton, New York.

Paterson AH. 2000. Comparative genomics of plant chromosomes. *The Plant Cell* **12**, 1523-1539.

Patterson HD and Thompson R. 1971. Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545-554.

Peacock L, Hunter T, Turner H and Brain P. 2001. Does host genotype diversity affect the distribution of insect and disease damage in willow cropping systems? *Journal of Applied Ecology* **38**, 1070-1081.

Pei MH, Royle D and Hunter T. 1993. Identity and host alternation of some willow rusts (*Melampsora* spp.) in England. *Mycological Research* **99**, 357-363.

Pei MH, Royle DJ and Hunter T. 1996. Pathogenic specialization in *Melampsora epitea* var. *epitea* on *Salix*. *Plant Pathology* **45**, 679-690.

Pei MH, Parker SE, Hunter T and Royle DJ. 1997. Variation in populations of *Melampsora* willow rust and the implications for design of short rotation coppice plantations. *Aspects of Applied Biology* **49**, 91-96.

Pei MH, Hunter T and Royle D. 1999. Host-pathogen relationship between *Salix* and *Melampsora* sheds light on the parentage of some biomass willows. *New Phytologist* **141**, 155-160.

Pei MH, Lindegaard KN, Hunter T and Ruiz C. 2002. Preliminary studies of inheritance of rust resistance in biomass willows. *Aspects of Applied Biology* **65**, 281-287.

Peng J, Carol P, Richards DE, King KE, Cowling RJ, Murphy GP and Harberd NP. 1997. The *Arabidopsis GAI* gene defines a signaling pathway that negatively regulates gibberrelin responses. *Genes and Development* **11**, 3194-3205.

Peng JR, Richards DE, Hartley NM, Murphy GP, Devos KM, Flintham JE, Beales J, Fish LK, Worland AJ, Pelica F, Sudhakar D, Christou P, Snape JW, Gale MD and

Harberd NP. 1999. 'Green revolution' genes encode mutant gibberellin response modulators. *Nature* **400**, 256-261.

Pflieger S, Lefebvre V and Causse M. 2001. The candidate gene approach in plant genetics: a review. *Molecular Breeding* **7**, 275-291.

Picoult-Newberg L, Idekar TE, Pohl MG, Taylor SL, Donaldson MA, Nickerson DA and Boyce-Jacino M. 1999. Mining SNPs from EST databases. *Genome Research* **9**, 167-174.

Pitcher KF and Hilton BS. 1997. ARBRE: from coppice establishment to gasification and generation of electricity. *Aspects of Applied Biology* **49**, 25-32.

Plomion C, Hurme P, Frigerio J-M, Ridolfi M, Pot M, Pionneau C, Avila C, Gallardo F, David H, Neutelings G, Campbell M, Canovas FM, Savolainen O, Bodénès C and Kremer A. 1999. Developing SSCP markers in two *Pinus* species. *Molecular Breeding* **5**, 21-31.

Powell W, Machray GC and Provan J. 1996. Polymorphism revealed by simple sequence repeats. *Trends in Plant Science* **1**, 215-222.

Rafalski A. 2002. Applications of single nucleotide polymorphism in crop genetics. *Current Opinion in Plant Biology* **5**, 94-100.

Rank NE, Köpf A, Julkunen-Tiitto R and Tahvanainen J. 1998. Host preference and larval performance of the salicylate-using leaf beetle *Phratora vitellinae*. *Ecology* **79**, 618-631.

Rebaï A, Goffinet B and Mangin B. 1994. Approximate thresholds of interval mapping tests for QTL detection. *Genetics* **138**, 235-240.

Ribaut J and Hoisington S. 1998. Marker-assisted selection: new tools and strategies. *Trends in Plant Science* **3**, 236-239.

Richmond T and Somerville D. 2000. Chasing the dream: plant EST microarrays. *Current Opinion in Plant Biology* **3**, 108-116.

Ritter E, Gebhardt C and Salamini F. 1990. Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between two heterozygous parents. *Genetics* **125**, 645-654.

Ritter E and Salamini F. 1996. Estimation of recombination frequencies in crosses of allogamous plant species with applications to linkage mapping. *Genetical Research*, **67**, 55-65.

Rivkin MI, Vallejos CE and McClean PE. 1999. Disease-resistance related sequences in common bean. *Genome* **42**, 41-47.

Robinson DL, Thompson R and Digby PGN. 1982. REML – a program for the analysis of non-orthogonal data by restricted maximum likelihood. *Compstat 1982 Proceedings in Computational Statistics, Part II (supplement)*, 231-232. Physica-Verlag, Vienna.

Robinson KM, Karp A and Taylor G. 2003. Defining leaf and canopy traits linked to high yield in short rotation coppice willow. *Biomass and Bioenergy* (In press).

Romagosa I, Han F, Ullrich SE, Hayes PM and Wesenberg DM. 1999. Verification of yield QTL through realized molecular marker-assisted selection responses in a barley cross.

Rönnberg-Wästljung AC and Gullberg U. 1999. Genetics of breeding characters with possible effects on biomass production in *Salix viminalis* (L.). *Theoretical and Applied Genetics* **98**, 531-540.

Rowell-Rahier M. 1984. The food plant preferences of *Phratora vitellinae* (Coleoptera: Chrysomelinae). B. A laboratory comparison of geographically isolated populations and experiments on conditioning. *Oecologia* **64**, 375-380.

Rowell-Rahier M and Pasteels JM. 1982. The significance of salicin for a *Salix*-feeder, *Phratora (Phyllodecta) vitellinae.* In: *Insect Plant Relationships*, (eds. Visser JH and Minks AK), Proceedings of the 5th international symposium, Insect-plant relationships. Sinnauer, Sunderland.

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning ZM, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schultz B, Cook L, Davenport R, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstiel B, Baldwin J, Stange-Thomann NS, Zody MC, Linton L, Lander ES and Altshuler D. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928-933.

Sage RB and Tucker K. 1997. Invertebrates in the canopy of willow and poplar short rotation coppices. *Aspects of Applied Biology* **49**, 105-120.

Saier MH. 1998. Genome sequencing and informatics: new tools for biochemical discoveries. *Plant Physiology* **117**, 1129-1133.

Saiki RK, Gelfand DH, Stoffel S, Scharf SS, Higuchi H, Horn GR, Mullis KB and Erlich HA. 1988. Primer directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487-491.

Sambrook J, Fritsch EF and Maniatis T (eds.). 1989. *Molecular Cloning: A Laboratory Manual.* 2nd Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Samils B, Lagercrantz U, Lascoux Ma and Gullberg U. 2001. Genetic structure of *Melampsora epitea* populations in Swedish *Salix viminalis* plantations. *European Journal of Plant Pathology* **107**, 399-409.

Sax K. 1923. The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris. Genetics* **8**, 552-560.

Schlötterer C and Tautz D. 1992. Slippage synthesis of simple sequence DNA. *Nucleic Acids Research* **20**, 211-215.

Schmidt R. 2002. Plant genome evolution: lessons from comparative genomics at the DNA level. *Plant Molecular Biology* **48**, 21-37.

Searle SR, Casella G and McCulloch CE. 1992. *Variance Components.* J.W. Wiley, New York.

Sebolt AM, Shoemaker RC and Diers BW. 2000. Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. *Crop Science* **40**, 1438-1444.

Shi MM. 2001. Enabling large-scale pharmacogenetic studies by high-throughput mutation detection and genotyping technologies. *Clinical Chemistry*, **47**, 164-172.

Shrimpton AE and Robertson A. 1988. Isolation of polygenic factors controlling bristle score in *Drosophila melanogaster.* 2. Distribution of third chromosome bristle effects within chromosome sections. *Genetics* **118**, 445-449.

Slocum MK, Figdore SS, Kennard WC, Suzuki JY and Osborn TC. 1990. Linkage arrangement of restriction fragment length polymorphisms in *Brassica oleracea. Theoretical and Applied Genetics* **80**, 57-64.

Skvortsov AK. 1968. Willows of the USSR. Proceedings of the study of the fauna and flora of the USSR. Published by the Moscow Society of Naturalists. New series, Botany section, Number 15 (XXIII).

Soetens P and Pasteels JM. 1994. Synergistic effect of secondary compounds and nutrients in the host plant choice of a salicaceous-feeding leaf beetle: *Phratora vitellinae* (Coleoptera: Chrysomelidae). *Mededelingen Faculteit Landbouwkundige en Toegepaste Biologische Wetenschappen Universiteit Gent* **59**, 685-689.

Sorrells ME and Anderson JA. 1996. Quantitative trait loci associated with preharvest sprouting in white wheat. In: Noda K and Mares DJ (Eds), *Preharvest sprouting in Cereals*, 137-142. Center for Academic Societies, Japan.

Sorrells ME and Wilson WA. 1997. Direct classification and selection of superior alleles for crop improvement. *Crop Science* **37**, 691-697.

Spickett SG, and Thoday JM. 1966. regular responses to selection. 3. interaction between located polygenes. *Genetical Research Cambridge* **7**, 96-121.

Stam P, Bos I, Qi X and Kiks R. 1997. QTL mapping of yield and yield-related traits in a wide cross of barley (*Hordeum vulgare*). In:, *Proceedings of the 10th Eucarpia Section Biometrics in Plant Breeding*, (ed. Krajewski P), 277-280. Poznan, Poland.

Stam P. 1998. Crop physiology, QTL analysis and plant breeding. In *Inherent variation in plant growth. Physiological mechanisms and ecological consequences*, 429-440. (Ed. Lmaber H, Poorter H and Van Vuuren MMI). Backhuys Publishers, Leiden, The Netherlands.

Stam P and Van Ooijen JW. 1995. Joinmap ™ version 2.0: Software for the calculation of genetic linkage maps. CPRO-DLO, Wageningen.

Staub JE, Kuhns LJ, May B and Grun P. 1982. Stability of potato tuber isozymes under different storage regimes. *Journal of American Science*, **107**, 405-408.

Staub JE, Serquen FC and Gupta M. 1996. Genetic markers, map construction, and their application in plant breeding. *HortScience* **31**, 729-741.

Sterky F, Regan S, Karlsson J, Hertzberg M, A Rhode, Holmberg A, Amini B, Bhalerao R, Larsson M, Villarroel R, Van Montagu M, Sandberg G, Olsson O, Teeri TT, Boerjan W, Gustafsson P, Uhlén M, Sundberg B and Lundeberg J. 1998. Gene discovery in the wood-forming tissues of poplar: Analysis of 5,692 expressed sequence tags. *Proceedings of the National Academy of Science, USA* **95**, 13330-13335.

Stott K. 1984. Improving the biomass potential of willow by selection and breeding. In: *Ecology and Management of Forest Biomass production systems.* **15**, 566-260. Ed. Perttu K. Department of Ecology and Environmental Research, Swedish University of Agricultural Sciences Report.

Stott K. 1992. Willows in the service of man. Willow symposium. Edinburgh 27-29 September 1991. *Proceedings of The Royal Society of Edinburgh* **98**

Suda Y and Argus GW. 1968. Chromosome numbers of some North American *Salix*. *Brittonia* **20**, 191-197.

Tahvanainen J, Julkunen-Tiitto R and Kettunen J. 1985. Phenolic glycosides govern the food selection pattern of willow feeding leaf beetles. *Oecologia* **67**, 52-56.

Tanksley SD, Young ND, Paterson AH and Bonierbale MW. 1989. RFLP mapping in plant breeding: new tools for an old science. *Biotechnology* **7**, 257-265.

Tanksley SD, Ganal MW, Prince JW, de Vincente MC, Bonierbale GB, Messeguer R, Miller JC, Miller L, Paterson AH, Pineda O, Röder MS, Wing RA, Wu W and Young ND. 1992. High density molecular linkage maps of the tomato and potato genomes. *Genetics* **132**, 1141-1160.

Tanksley SD, Ganal MW and Martin GB. 1995. Chromosome landing: a paradigm for map-based cloning in plants with large genomes. *Trends in Genetics* **11**, 63-68.

Tanksley SD and McCouch SR. 1997. Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* **277**, 1063-1066.

Tautz D. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research* **17**, 6463-6471.

Temesgen B, Brown GR, Harry DE, Kinlaw CS, Sewell MM and Neale DB. 2001. Genetic mapping of expressed sequence tag polymorphism (ESTP) markers in loblolly pine (*Pinus taeda* L.). *Theoretical and Applied Genetics* **102**, 664-675.

Thibault J. 1998. Nuclear DNA amount in pure species and hybrid willows (*Salix*): a flow cytometric investigation. *Canadian Journal of Botany* **76**, 157-165.

Thieme H. 1971. Vorkommen and verbreitung von phenolglykosiden den in der familie der Salicaceen. *Herba Pol* **17**, 248-257.

Thoday JM. 1961. Location of polygenes. *Nature* **191**, 368-370.

Tong X and Smith LM. 1992. Solid phase method for the purification of DNA sequencing reactions. *Analytical Chemistry* **64**, 2672-2677.

Tragoonrung S, Kanazin V, Hayes PM, Blake TM. 1992. Sequence-tagged-site-facilitated PCR for barley genome mapping. *Theoretical and Applied Genetics* **94**, 764-772.

Triest L, DeGreef B, DeBondt R, VandenBossche D and DHaeseleer M. 1997. Use of RAPD markers to estimate hybridization in *Salix alba* and *Salix fragilis*. *Belgian Journal of Botany* **129**, 140-148.

Tsarouhas V, Gullberg U and Lagercrantz U. 2001. An AFLP and RFLP linkage map and quantitative trait locus (QTL) analysis of growth traits in *Salix*. *Theoretical and Applied Genetics* **105**, 277-288.

Tucker K and Sage R. 1999. Integrated pest management in short rotation coppice for energy – *a growers guide*. Game conservancy Ltd, Hampshire, UK.

Turnbull DJ. 1997. The commercial reality of producing biomass from short rotation coppice. *Aspects of Applied Biology* **49**, 71-78.

Van Hintum TJL and Van Treuren R. 2002. Molecular markers: Tools to rimporve genbank efficiency. *Cellular and Molecular Biology Letters* **7**, 737-744.

Van Ooijen J, Sandbrink H, Purimahua C, Vrielink R, Verkerk R, Zabel P and Lindhout P. 1993. Mapping quantitative genes involved in a trait assessed on an ordinal scale: A case study with bacterial canker in *Lycopersicon peruvianum*. In:. *Molecular Biology of the Tomato*. (ed. Yoder JI), 59-74. Technomic Publishing Co. Lancaster PA.

Van Ooijen JW. 1999. LOD significance thresholds for QTL analysis in experimental populations of diploid species. *Heredity* **83**, 613-324.

Van Ooijen JW. 1992. Accuracy of mapping quantitative trait loci in autogamous species. *Theoretical and Applied Genetics* **804**, 803-811.

Van Ooijen JW, Boer MP, Jansen RC and Maliepaard C. 2002. MapQTL® 4.0, Software for the calculation of QTL positions on genetic maps. Plant Research International, Wageningen, the Netherlands.

Van Ooijen JW and Voorips RE. 2001. JoinMap® 3.0, Software for the calculation of genetic linkage maps. Plant Research International, Wageningen, The Netherlands.

Verhaegen D and Plomion C. 1996. Genetic mapping of *Eucalyptus urophylla* and *Eucalyptus grandis* using RAPD markers. *Genome* **39**: 1051-1061.

Viherä-Aarnio A. 1988. Variation in coppicing characteristics and biomass production of willow clones and hybrids from the breeders point of view. In: *Proceedings of the IEA Workshop on Cell Culture and Coppicing*, 101-112.

Voorips RE. 2002. MapChart: Software for the graphical presentation of linkage maps and QTLs. *Journal of Heredity* **93,** 77-78.

Vos P, Hogers R, Bleeker M, Reijans M, Vandelee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M and Zabeau M. 1995. AFLP- A new technique for DNA fingerprinting. *Nucleic Acids Research* **23,** 4407-4414.

Vos P and Kuiper M. 1997. AFLP analysis. In: *DNA markers: protocols, applications and overviews.* (eds. Caetano-Anolles G and Gresshoff PM). Wiley-Liss Inc., New York, 115-131.

Weber JL. 1990. Informativeness of human (DC-DA)n. (DG-DT)n polymorphisms. *Genomics* **7,** 524-530.

Welsh J and McClelland M. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research* **18,** 7213-7218.

Williams GK, Kubelik AR, Livak KJ, Rafalski JA and Tingey SV. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* **18,** 6531-6535.

Wiltshire CW, Kendall DA, Hunter T and Arnold GM. 1997. Host-plant preferences of two willow-feeding leaf beetles (Coleoptera, Chrysomelidae). *Aspects of Applied Biology* **49,** 113-120.

Wiren A and Larsson S. 1984. Preferences of insects for different willow clones: a case study with *Galerucella lineola* (Col. Chrysomelidae). In *Ecology and management of forest biomass production systems. Report Number 15.* (ed. Perttu K).

Department of Ecology and Environmental Research, Swedish University of Agricultural Sciences, Uppsala, 383-389.

Wrather JA. 1998. Soybean disease loss estimates in the U.S.A. National Soybean Breeders Workshop, St Louis, MO, American Soybean Association.

Wright S. 1934. An analysis of variability in number of digits in an inbred strain of guinea pigs. *Genetics* **19**, 506-536.

Wu RL. 1998. Genetic mapping of QTLs affecting tree growth and architecture in *Populus*: implication for ideotype breeding. *Theoretical and Applied Genetics* **96**, 447-457.

Wu RL, Han YF, Hu JJ, Fang JJ, Li L, Li ML and Zeng ZB. 2000. An integrated genetic map of *Populus deltoides* based on amplified fragment length polymorphisms. *Theoretical and Applied Genetics* **100**, 1249-1256.

Wyman AR and White R. 1980. A highly polymorphic locus in human DNA. *Proceedings of the National Academy of Sciences USA* **77**, 6754-6758.

Young ND. 1996. QTL mapping and quantitative disease resistance in plants. *Annual Review of Phytopathology* **34**, 479-501.

Young ND. 1999. A cautiously optimistic vision for marker-assisted breeding. *Molecular Breeding* **5**, 505-510.

Young,WP, Schupp JM and Keim P. 1999. DNA methylation and AFLP distribution in the soybean genome. *Theoretical and Applied Genetics* **99**, 785-792.

Zabeau M and Vos P. 1993. Selective restriction fragment amplification, a general method for DNA fingerprinting. European Patent Application, publication no: EP 0534858-A1, No. 92402629.7.

Appendix I. Experimental design for the 2000 laboratory-based beetle feeding experiments based on mapping populations K1, K8 and K3 (Section 4.2.7.2).

Population K1

K1	Position in dish										Position in dish									
Dish	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	Dish	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
1	699	352	751	865	29	216	308	201	633	142	47	840	865	669	475	78	52	229	743	451	424
2	352	89	702	664	81	374	686	806	868	52	48	299	664	475	115	633	235	168	658	786	324
3	89	305	840	252	256	205	P2	244	142	235	49	865	252	115	171	868	420	164	159	518	502
4	305	133	299	353	216	919	116	199	52	420	50	664	353	171	680	142	819	743	939	320	265
5	133	929	865	788	374	683	614	410	235	819	51	699	252	788	680	615	52	82	658	654	874
6	929	751	664	433	205	877	538	669	420	82	52	352	353	433	615	78	235	33	159	236	425
7	751	702	252	555	919	488	715	475	819	33	53	89	788	555	78	633	420	25	939	309	225
8	702	840	353	93	683	300	201	115	82	25	54	305	433	93	633	868	819	652	654	463	424
9	840	299	788	860	877	308	806	171	33	652	55	133	555	860	868	142	82	820	236	451	324
10	299	865	433	328	488	686	244	680	25	820	56	929	93	328	142	52	33	776	309	786	502
11	865	664	555	P1	300	P2	199	615	652	776	57	751	860	P1	52	235	25	603	463	518	265
12	664	252	93	177	308	116	410	78	820	603	58	699	702	328	177	235	420	652	74	451	320
13	252	353	860	607	686	614	669	633	776	74	59	352	840	P1	607	420	819	820	554	786	874
14	353	788	328	29	P2	538	475	868	603	554	60	89	299	177	29	819	82	776	229	518	425
15	788	433	P1	81	116	715	115	142	74	229	61	305	865	607	81	82	33	603	168	320	225
16	433	555	177	256	614	201	171	52	554	168	62	133	664	29	256	33	25	74	164	874	424
17	555	93	607	216	538	806	680	235	229	164	63	929	252	81	216	25	652	554	743	425	324
18	93	860	29	374	715	244	615	420	168	743	64	751	353	256	374	652	820	229	658	225	502
19	860	328	81	205	201	199	78	819	164	658	65	702	788	216	205	820	776	168	159	424	265
20	328	P1	256	919	806	410	633	82	743	159	66	699	840	433	374	919	776	603	164	939	324
21	P1	177	216	683	244	669	868	33	658	939	67	352	299	555	205	683	603	74	743	654	502
22	177	607	374	877	199	475	142	25	159	654	68	89	865	93	919	877	74	554	658	236	265
23	607	29	205	488	410	115	52	652	939	236	69	699	305	664	860	683	488	554	229	159	309
25	29	81	919	300	669	171	235	820	654	309	70	352	133	252	328	877	300	229	168	939	463
25	81	256	683	308	475	680	420	776	236	463	71	89	929	353	P1	488	308	168	164	654	451
26	256	216	877	686	115	615	819	603	309	451	72	305	751	788	177	300	686	164	743	236	786
27	216	374	488	P2	171	78	82	74	463	786	73	133	702	433	607	308	P2	743	658	309	518
28	374	205	300	116	680	633	33	554	451	518	74	929	840	555	29	686	116	658	159	463	320
29	205	919	308	614	615	868	25	229	786	320	75	751	299	93	81	P2	614	159	939	451	874
30	919	683	686	538	78	142	652	168	518	874	76	702	865	860	256	116	538	939	654	786	425
31	683	877	P2	715	633	52	820	164	320	425	77	840	664	328	216	614	715	654	236	518	225
32	877	488	116	201	868	235	776	743	874	225	78	299	252	P1	374	538	201	236	309	320	424
33	488	300	614	806	142	420	603	658	425	424	79	865	353	177	205	715	806	309	463	874	324
34	300	308	538	244	52	819	74	159	225	324	80	664	788	607	919	201	244	463	451	425	502
35	308	686	715	199	235	82	554	939	424	502	81	252	433	29	683	806	199	451	786	225	265
36	686	P2	201	410	420	33	229	654	324	265	82	699	353	555	81	877	244	410	786	518	424
37	699	P2	116	806	669	819	25	168	236	502	83	352	788	93	256	488	199	669	518	320	324
38	352	116	614	244	475	82	652	164	309	265	84	89	433	860	216	300	410	475	320	874	502
39	699	89	614	538	199	115	33	820	743	463	85	305	555	328	374	308	669	115	874	425	265
40	352	305	538	715	410	171	25	776	658	451	86	699	133	93	P1	205	686	475	171	425	225
41	89	133	715	201	669	680	652	603	159	786	87	352	929	860	177	919	P2	115	680	225	424
42	305	929	201	806	475	615	820	74	939	518	88	89	751	328	607	683	116	171	615	424	324
43	133	751	806	244	115	78	776	554	654	320	89	305	702	P1	29	877	614	680	78	324	502
44	929	702	244	199	171	633	603	229	236	874	90	133	840	177	81	488	538	615	633	502	265
45	751	840	199	410	680	868	74	168	309	425	91	699	929	299	607	256	300	715	78	868	265
46																					

Population K8

	Positon in dish											*Position in dish*									
Dish	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	***Dish***	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
1	597	50	230	905	547	347	920	84	909	19	**47**	256	905	728	421	440	786	297	287	507	10
2	50	695	315	94	71	192	350	422	P2	786	**48**	309	94	421	241	909	319	P1	880	648	767
3	695	292	256	106	183	503	386	931	19	319	**49**	905	106	241	176	P2	436	420	517	774	583
4	292	781	309	381	347	166	323	825	786	436	**50**	94	381	176	197	19	434	287	127	581	866
5	781	614	905	460	192	628	470	805	319	434	**51**	597	106	460	197	732	786	180	880	295	77
6	614	230	94	763	503	921	759	728	436	180	**52**	50	381	763	732	440	319	868	517	46	326
7	230	315	106	520	166	132	548	421	434	868	**53**	695	460	520	440	909	436	379	127	542	608
8	315	256	381	817	628	438	84	241	180	379	**54**	292	763	817	909	P2	434	359	295	912	10
10	256	309	460	555	921	920	422	176	868	359	**55**	781	520	555	P2	19	180	189	46	507	767
10	309	905	763	414	132	350	931	197	379	189	**56**	614	817	414	19	786	868	939	542	648	583
11	905	94	520	841	438	386	825	732	359	939	**57**	230	555	841	786	319	379	915	912	774	866
12	94	106	817	237	920	323	805	440	189	915	**58**	597	315	414	237	319	436	359	577	507	581
13	106	381	555	26	350	470	728	909	939	577	**59**	50	256	841	26	436	434	189	772	648	77
14	381	460	414	547	386	759	421	P2	915	772	**60**	695	309	237	547	434	180	939	297	774	326
15	460	763	841	71	323	548	241	19	577	297	**61**	292	905	26	71	180	868	915	P1	581	608
16	763	520	237	183	470	84	176	786	772	P1	**62**	781	94	547	183	868	379	577	420	77	10
17	520	817	26	347	759	422	197	319	297	420	**63**	614	106	71	347	379	359	772	287	326	767
18	817	555	547	192	548	931	732	436	P1	287	**64**	230	381	183	192	359	189	297	880	608	583
19	555	414	71	503	84	825	440	434	420	880	**65**	315	460	347	503	189	939	P1	517	10	866
20	414	841	183	166	422	805	909	180	287	517	**66**	597	256	763	192	166	939	915	420	127	767
21	841	237	347	628	931	728	P2	868	880	127	**67**	50	309	520	503	628	915	577	287	295	583
22	237	26	192	921	825	421	19	379	517	295	**68**	695	905	817	166	921	577	772	880	46	866
23	26	547	503	132	805	241	786	359	127	46	**69**	597	292	94	555	628	132	772	297	517	542
24	547	71	166	438	728	176	319	189	295	542	**70**	50	781	106	414	921	438	297	P1	127	912
26	71	183	628	920	421	197	436	939	46	912	**71**	695	614	381	841	132	920	P1	420	295	507
26	183	347	921	350	241	732	434	915	542	507	**72**	292	230	460	237	438	350	420	287	46	648
27	347	192	132	386	176	440	180	577	912	648	**73**	781	315	763	26	920	386	287	880	542	774
28	192	503	438	323	197	909	868	772	507	774	**74**	614	256	520	547	350	323	880	517	912	581
29	503	166	920	470	732	P2	379	297	648	581	**75**	230	309	817	71	386	470	517	127	507	77
30	166	628	350	759	440	19	359	P1	774	77	**76**	315	905	555	183	323	759	127	295	648	326
31	628	921	386	548	909	786	189	420	581	326	**77**	256	94	414	347	470	548	295	46	774	608
32	921	132	323	84	P2	319	939	287	77	608	**78**	309	106	841	192	759	84	46	542	581	10
33	132	438	470	422	19	436	915	880	326	10	**79**	905	381	237	503	548	422	542	912	77	767
34	438	920	759	931	786	434	577	517	608	767	**80**	94	460	26	166	84	931	912	507	326	583
35	920	350	548	825	319	180	772	127	10	583	**81**	106	763	547	628	422	825	507	648	608	866
36	350	386	84	805	436	868	297	295	767	866	**82**	597	381	520	71	921	931	805	648	774	10
37	597	386	323	422	728	434	379	P1	46	583	**83**	50	460	817	183	132	825	728	774	581	767
38	50	323	470	931	421	180	359	420	542	866	**84**	695	763	555	347	438	805	421	581	77	583
39	597	695	470	759	825	241	868	189	287	912	**85**	292	520	414	192	920	728	241	77	326	866
40	50	292	759	548	805	176	379	939	880	507	**86**	597	781	817	841	503	350	421	176	326	608
41	695	781	548	84	728	197	359	915	517	648	**87**	50	614	555	237	166	386	241	197	608	10
42	292	614	84	422	421	732	189	577	127	774	**88**	695	230	414	26	628	323	176	732	10	767
43	781	230	422	931	241	440	939	772	295	581	**89**	292	315	841	547	921	470	197	440	767	583
44	614	315	931	825	176	909	915	297	46	77	**90**	781	256	237	71	132	759	732	909	583	866
45	230	256	825	805	197	P2	577	P1	542	326	**91**	597	614	309	26	183	438	548	440	P2	866
46	315	309	805	728	732	19	772	420	912	608											

Population K3

	Position in dish									
Dish	1	2	3	4	5	6	7	8	9	10
1	48	4	38	158	P1	101	44	128	114	19
2	4	8	18	2	188	164	60	131	29	74
3	8	66	15	47	31	21	14	1	19	P2
4	66	75	45	10	101	118	180	49	74	12
5	75	88	158	5	164	63	46	40	P2	55
6	88	38	2	146	21	62	9	35	12	33
7	38	18	47	65	118	26	59	52	55	22
8	18	15	10	20	63	16	128	51	33	37
9	15	45	5	25	62	44	131	53	22	34
10	45	158	146	39	26	60	1	96	37	27
11	158	2	65	50	16	14	49	57	34	41
12	2	47	20	43	44	180	40	7	27	3
13	47	10	25	42	60	46	35	114	41	132
14	10	5	39	P1	14	9	52	29	3	56
15	5	146	50	188	180	59	51	19	132	24
16	146	65	43	31	46	128	53	74	56	32
17	65	20	42	101	9	131	96	P2	24	61
18	20	25	P1	164	59	1	57	12	32	203
19	25	39	188	21	128	49	7	55	61	98
20	39	50	31	118	131	40	114	33	203	6
21	50	43	101	63	1	35	29	22	98	23
22	43	42	164	62	49	52	19	37	6	17
23	42	P1	21	26	40	51	74	34	23	142
24	P1	188	118	16	35	53	P2	27	17	204
25	188	31	63	44	52	96	12	41	142	58
26	31	101	62	60	51	57	55	3	204	28
27	101	164	26	14	53	7	33	132	58	136
28	164	21	16	180	96	114	22	56	28	30
29	21	118	44	46	57	29	37	24	136	100
30	118	63	60	9	7	19	34	32	30	36
31	63	62	14	59	114	74	27	61	100	11
32	62	26	180	128	29	P2	41	203	36	13
33	26	16	46	131	19	12	3	98	11	108
34	16	44	9	1	74	55	132	6	13	175
35	44	60	59	49	P2	33	56	23	108	64
36	60	14	128	40	12	22	24	17	175	54
37	48	14	180	131	35	55	37	32	142	64
38	4	180	46	1	52	33	34	61	204	54
39	48	8	46	9	49	51	22	27	203	58
40	4	66	9	59	40	53	37	41	98	28
41	8	75	59	128	35	96	34	3	6	136
42	66	88	128	131	52	57	27	132	23	30
43	75	38	131	1	51	7	41	56	17	100
44	88	18	1	49	53	114	3	24	142	36
45	38	15	49	40	96	29	132	32	204	11
46	18	45	40	35	57	19	56	61	58	13
47	15	158	35	52	7	74	24	203	28	108
48	45	2	52	51	114	P2	32	98	136	175
49	158	47	51	53	29	12	61	6	30	64
50	2	10	53	96	19	55	203	23	100	54
51	48	47	5	96	57	74	33	98	17	36
52	4	10	146	57	7	P2	22	6	142	11
53	8	5	65	7	114	12	37	23	204	13
54	66	146	20	114	29	55	34	17	58	108
55	75	65	25	29	19	33	27	142	28	175
56	88	20	39	19	74	22	41	204	136	64
57	38	25	50	74	P2	37	3	58	30	54
58	48	18	39	43	P2	12	34	132	28	100
59	4	15	50	42	12	55	27	56	136	36
60	8	45	43	P1	55	33	41	24	30	11
61	66	158	42	188	33	22	3	32	100	13
62	75	2	P1	31	22	37	132	61	36	108
63	88	47	188	101	37	34	56	203	11	175
64	38	10	31	164	34	27	24	98	13	64
65	18	5	101	21	27	41	32	6	108	54
66	48	15	146	164	118	41	3	61	23	175
67	4	45	65	21	63	3	132	203	17	64
68	8	158	20	118	62	132	56	98	142	54
69	48	66	2	25	63	26	56	24	6	204
70	4	75	47	39	62	16	24	32	23	58
71	8	88	10	50	26	44	32	61	17	28
72	66	38	5	43	16	60	61	203	142	136
73	75	18	146	42	44	14	203	98	204	30
74	88	15	65	P1	60	180	98	6	58	100
75	38	45	20	188	14	46	6	23	28	36
76	18	158	25	31	180	9	23	17	136	11
77	15	2	39	101	46	59	17	142	30	13
78	45	47	50	164	9	128	142	204	100	108
79	158	10	43	21	59	131	204	58	36	175
80	2	5	42	118	128	1	58	28	11	64
81	47	146	P1	63	131	49	28	136	13	54
82	48	10	65	188	62	1	40	136	30	108
83	4	5	20	31	26	49	35	30	100	175
84	8	146	25	101	16	40	52	100	36	64
85	66	65	39	164	44	35	51	36	11	54
86	48	75	20	50	21	60	52	53	11	13
87	4	88	25	43	118	14	51	96	13	108
88	8	38	39	42	63	180	53	57	108	175
88	66	18	50	P1	62	46	96	7	175	64
90	75	15	43	188	26	9	57	114	64	54
91	48	88	45	42	31	16	59	7	29	54

Appendix II. Experimental design for the 2001 laboratory-based beetle feeding experiments based on the K8 mapping population (Section 6.2.1.4).

Population K8

Rep	Dish	1	2	3	4	5	6	7	8	9	10	11
		Position in dish										
1	1	37	86	33	18	71	3	90	52	105	120	56
	2	45	75	26	7	22	113	109	60	94	41	79
	3	19	23	57	4	38	121	53	106	87	91	72
	4	119	17	70	36	51	32	66	85	104	89	2
	5	73	58	92	20	24	88	54	107	111	5	39
	6	21	59	40	74	108	112	78	25	93	55	6
	7	118	69	35	103	16	31	50	84	99	65	1
	8	116	48	97	101	29	67	10	44	14	63	82
	9	110	42	80	12	95	61	8	114	46	76	27
	10	102	117	15	11	34	64	83	68	49	98	30
	11	77	43	96	13	81	9	115	100	62	47	28
2	12	96	79	21	50	4	119	56	27	102	44	73
	13	101	49	72	66	118	95	78	20	26	43	3
	14	63	86	40	69	115	11	109	46	23	17	92
	15	57	51	97	74	80	103	22	5	28	34	120
	16	15	55	61	32	90	38	9	107	84	113	67
	17	18	47	70	110	1	41	93	64	87	116	24
	18	58	35	81	98	121	75	52	104	6	29	12
	19	36	82	99	111	76	105	7	53	30	13	59
	20	16	39	91	33	45	108	68	85	62	114	10
	21	77	83	8	112	37	14	89	54	106	31	60
	22	65	71	42	88	19	2	117	100	48	25	94
3	23	41	55	58	27	13	103	89	10	72	86	117
	24	39	8	84	22	56	25	53	98	101	70	115
	25	47	92	106	30	78	61	44	2	16	75	120
	26	77	18	80	108	111	49	32	63	94	35	4
	27	51	23	110	6	96	65	20	37	82	68	113
	28	116	85	54	102	9	57	40	26	99	71	12
	29	91	74	43	105	88	15	119	60	1	29	46
	30	76	93	34	17	62	79	31	48	3	121	107
	31	97	66	38	52	7	24	114	83	100	21	69
	32	73	104	28	59	14	118	45	42	87	90	11
	33	50	95	19	109	64	5	112	81	33	36	67
4	34	16	27	38	60	49	71	104	93	82	5	115
	35	85	96	8	41	52	107	19	118	74	63	30
	36	72	50	116	17	28	61	94	105	39	83	6
	37	89	111	45	12	78	67	1	23	56	34	100
	38	54	21	98	109	120	65	10	43	32	87	76
	39	20	64	86	31	42	9	75	119	97	108	53
	40	36	91	69	80	14	102	25	3	58	113	47
	41	11	44	22	33	77	121	55	66	99	110	88
	42	114	48	103	26	4	70	15	92	81	59	37
	43	79	35	101	24	13	68	46	112	2	57	90
	44	29	84	18	117	73	7	51	106	95	40	62
5	45	48	112	105	41	73	80	9	23	16	66	98
	46	62	12	37	5	119	55	69	87	30	94	101
	47	116	27	2	77	84	20	91	34	59	52	109
	48	86	118	68	22	61	93	29	100	54	36	4
	49	8	15	47	104	65	33	79	97	111	72	40
	50	96	64	78	121	32	39	14	71	103	46	7
	51	28	21	110	35	92	60	53	67	117	3	85
	52	11	82	25	89	107	75	18	114	50	43	57
	53	10	49	113	106	24	81	42	99	56	17	74
	54	31	38	70	45	63	102	88	6	120	95	13
	55	51	44	108	1	115	26	58	76	90	19	83
6	56	118	39	23	76	2	81	60	18	102	55	97
	57	28	112	70	12	7	107	91	44	65	86	49
	58	29	8	108	92	113	50	71	66	87	13	34
	59	42	21	26	105	121	63	5	89	84	68	47
	60	35	9	88	93	114	72	14	56	30	109	51
	61	64	111	106	48	22	6	43	69	85	27	90
	62	99	62	41	25	46	4	20	67	104	83	120
	63	33	96	54	101	38	1	59	17	75	80	117
	64	32	100	58	95	74	16	116	37	79	11	53
	65	52	36	57	73	31	10	15	110	115	94	78
	66	77	119	24	3	19	61	103	98	40	45	82

Rep	Dish	Position in dish 1	2	3	4	5	6	7	8	9	10	11
7	67	101	16	88	113	4	40	76	64	28	52	89
	68	87	75	15	99	39	63	27	51	3	112	100
	69	7	56	104	43	19	31	92	80	55	68	116
	70	41	5	102	17	29	78	114	53	90	65	77
	71	109	1	13	25	73	97	37	85	49	121	61
	72	110	50	86	14	74	62	111	38	26	2	98
	73	54	18	115	42	30	6	91	67	103	66	79
	74	58	34	21	106	82	70	118	33	94	46	9
	75	96	11	120	84	108	36	72	48	60	12	24
	76	47	22	59	95	119	23	107	71	10	35	83
	77	45	32	117	20	57	8	93	44	105	69	81
8	78	96	91	89	98	97	93	90	99	92	95	94
	79	43	35	38	39	37	44	41	36	34	40	42
	80	16	15	22	21	19	14	20	18	17	13	12
	81	6	4	1	7	5	2	10	8	3	11	9
	82	114	113	118	112	111	116	119	121	120	115	117
	83	88	82	84	87	83	85	80	86	81	79	78
	84	103	109	104	110	100	108	107	106	105	102	101
	85	50	51	47	55	49	52	46	54	48	53	45
	86	61	63	59	56	60	57	64	62	58	65	66
	87	75	74	67	73	76	68	70	77	69	71	72
	88	30	28	27	31	32	29	25	23	24	33	26
9	89	67	121	102	94	24	16	51	59	8	86	43
	90	47	74	117	4	39	82	90	66	31	109	12
	91	53	34	10	112	26	104	96	69	88	18	61
	92	108	46	73	65	38	30	81	89	3	22	116
	93	45	72	29	99	21	2	64	37	107	115	80
	94	70	54	62	78	27	35	11	113	97	19	105
	95	50	101	15	120	23	77	7	93	58	42	85
	96	92	100	41	57	49	33	6	84	119	14	76
	97	40	13	32	48	91	110	75	56	118	83	5
	98	9	111	52	44	103	60	17	25	87	95	68
	99	28	114	36	79	55	98	63	71	20	106	1
10	100	37	27	47	17	7	67	57	98	88	108	118
	101	56	87	16	117	46	36	107	77	97	26	6
	102	110	49	19	29	89	120	9	79	59	69	39
	103	55	85	65	24	115	4	34	75	14	105	95
	104	45	5	25	116	15	66	86	35	106	76	96
	105	99	109	68	28	48	18	8	119	58	78	38
	106	10	30	80	40	70	60	90	100	50	20	121
	107	32	1	82	72	92	112	42	22	102	52	62
	108	41	51	91	101	81	61	11	31	71	21	111
	109	13	44	104	3	74	94	84	114	54	23	64
	110	33	83	12	73	2	43	93	53	63	103	113
11	111	51	33	13	69	107	98	42	116	78	60	4
	112	87	2	31	96	58	40	67	105	49	22	114
	113	86	57	1	48	21	30	77	39	113	95	104
	114	119	34	16	72	7	110	63	81	25	54	90
	115	91	82	64	8	35	17	100	26	55	73	120
	116	70	117	23	5	43	52	79	108	14	61	99
	117	11	20	103	38	47	94	56	112	29	85	76
	118	15	24	89	80	44	53	6	71	109	62	118
	119	93	75	10	46	28	37	84	66	19	111	102
	120	101	27	65	92	9	45	83	18	121	36	74
	121	50	88	97	32	106	3	12	59	115	41	68
12	122	59	31	100	18	72	5	98	85	113	46	44
	123	109	42	55	70	111	83	29	16	96	57	3
	124	74	33	115	7	48	89	102	20	87	35	61
	125	30	84	45	17	4	112	43	71	58	110	97
	126	13	93	65	80	52	11	39	26	106	119	67
	127	78	24	63	9	22	117	37	50	91	104	76
	128	95	54	121	2	56	15	28	41	82	69	108
	129	105	92	118	25	64	12	51	79	10	77	38
	130	68	66	40	14	81	107	27	53	94	1	120
	131	116	62	23	75	36	88	49	8	103	90	21
	132	34	86	73	101	32	47	60	6	114	99	19

UNIVERSITY